UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR

PROYECTO FIN DE CARRERA

Ingeniería de Telecomunicación

EVALUACIÓN COMPARATIVA DE TÉCNICAS DE DETECCIÓN Y DESCRIPCIÓN DE PUNTOS DE INTERÉS EN IMÁGENES

Miguel Martín Redondo

Julio 2016

EVALUACIÓN COMPARATIVA DE TÉCNICAS DE DETECCIÓN Y DESCRIPCIÓN DE PUNTOS DE INTERÉS EN IMÁGENES

AUTOR: Miguel Martín Redondo
TUTOR: Fulgencio Navarro Fajardo
PONENTE: Jesús Bescós Cano

Video Processing and Understanding Lab
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio 2016

Trabajo parcialmente financiado por el gobierno español bajo el proyecto TEC2014-53176-R (HA-Video)

Resumen.

Uno de los objetivos principales del proyecto ha sido elaborar un marco de evaluación de algoritmos de detección y descripción de puntos de interés que permitiese actualizar las referencias del estado del arte en este sentido.

En base a dicho marco de evaluación propuesto, se ha definido el otro de los objetivos principales del proyecto, la realización de una evaluación comparativa de las técnicas de puntos de interés más relevantes en el estado del arte en la actualidad.

Para todo ello, se ha trabajado en tres grandes bloques.

En primer lugar, a partir de un análisis de las fortalezas y debilidades de los marcos de evaluación propuestos con anterioridad en el estado del arte, se propuso un nuevo conjunto de datos, acompañado de una metodología y métricas de evaluación. Estas propuestas permitieron definir un nuevo marco de evaluación sobre el que se trabajaría en el resto del proyecto.

En segundo lugar, se realizó un exhaustivo estudio del estado del arte, que llevó a categorizar y seleccionar una serie de algoritmos de puntos de interés que se consideraron como los más relevantes en la actualidad.

Sobre este conjunto de algoritmos, se desarrolló el tercer y último bloque, que constó de una extensa evaluación comparativa de todos ellos sobre el marco de evaluación propuesto anteriormente.

Con ello, se considera que a la finalización del proyecto, se ha logrado alcanzar ambos objetivos principales, la propuesta de un nuevo marco de evaluación, y la realización de una evaluación comparativa que pueda ser referencia para la comunidad científica.

Palabras clave

Puntos de interés, detección, descripción, evaluación comparativa, marco de evaluación.

Abstract.

One of the main objectives of this project has been to develop a framework for the evaluation of keypoint detection and description algorithms, in order to update the references of the state of the art in this topic.

Based on this framework, the other main objective has been to perform a comparative evaluation of the state of the art algorithms in local features field.

To this aim, three main stages have been faced.

The first one has been to propose a new dataset, together with a evaluation methodology, based on an analysis of the strengths and weaknesses of previous frameworks in the state of the art. This proposal will set a new evaluation framework for the following stages.

The second one, an exhaustive study of the state of the art allowed to select the main techniques and categorize them according to its properties.

Finally, those selected techniques were tested on the proposed evaluation framework.

Summarizing, at the end of the project, both objectives, the evaluation framework proposal, and the comparative evaluation, have been satisfied.

Key words

Keypoints, detection, description, comparative evaluation, evaluation framework.

Agradecimientos.

Quiero agradecer en primer lugar a mi tutor, Fulgencio Navarro, por su ayuda y su apoyo para la consecución de este proyecto. Haber podido desarrollarlo y trabajar junto a él ha sido una gran experiencia y aprendizaje que me llevo conmigo.

Quiero agradecer también al resto de integrantes del VPULab por la gran predisposición que han mostrado cuando he necesitado ayuda y por el buen ambiente que crean en el laboratorio. Igualmente quiero agradecer a los jefes del grupo, José María Martínez y Jesús Bescós por haberme brindado la oportunidad de llevar a cabo este proyecto.

No puedo dejar de nombrar en estos agradecimientos a mi amigo Alberto, compañero de viaje todos estos años.

Y por supuesto, sin el apoyo de mi familia no habría llegado hasta aquí. A mi padre y a mi madre, esto es por vosotros, muchas gracias.

Gracias a todos.

Miguel Martín Redondo.

Junio 2016.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción.

1.1. Motivación.

Los puntos de interés son la base de numerosas tareas de la visión por computador. La detección y descripción de puntos de interés, dentro del marco del procesado de imagen, son la base de tareas tan diversas como reconocimiento de objetos, *tracking* o modelado 3D. Sus notables resultados han motivado a la comunidad científica a trabajar en esta dirección, produciéndose un crecimiento considerable de mejoras y nuevas técnicas propuestas en los últimos años.

No obstante, pese a la cantidad de algoritmos que han ido surgiendo, no han aparecido nuevos estudios comparativos de referencia en el estado del arte. La ausencia de los mismos, en conjunto con el elevado número de propuestas, hace que para el investigador que recurre a estas técnicas para sus aplicaciones, la elección de la idónea sea una tarea compleja y no siempre fructífera.

Por último, otro aspecto que complica la comparación entre las técnicas de puntos de interés del estado del arte es la heterogeneidad de los marcos de evaluación, principalmente a nivel de métricas-*dataset* (conjunto de datos). Si bien es cierto que existe algún *dataset* de referencia, la técnicas que lo utilizan comienza a sobre ajustarse. Aquellas técnicas que se comparan sobre otros conjuntos de datos se justifican adicionalmente en la elevada complejidad de algunas tareas y la nimiedad de otras.

1.2. Objetivos.

El objetivo principal de este proyecto será, por lo tanto, proporcionar a la comunidad científica una comparativa actualizada de las técnicas de puntos de interés del estado del arte. Habida cuenta de la complejidad de la tarea, así como de los pro-

blemas presentados por los actuales marcos de evaluación, el objetivo principal del proyecto se abordará desde la consecución de una serie de objetivos parciales.

- Se llevará a cabo un análisis de los estudios comparativos existentes en el estado del arte, extrayendo fortalezas y debilidades que faciliten la propuesta.
- Se diseñará, grabará y anotará un nuevo y exhaustivo conjunto de datos de evaluación. Dicho conjunto presentará todas las propiedades de interés que deben presentar estas técnicas, así como numerosas combinaciones de las mismas. En conjunto con los datos, se definirán una métricas y metodologías comunes de evaluación.
- Se analizarán las técnicas de puntos de interés del estado del arte dividiéndolas en sus dos etapas principales, la detección y la descripción.
- Una vez analizadas, se categorizarán y se seleccionarán aquellas más punteras en cada categoría de cara a su posterior evaluación. La selección de los métodos se acompañará de un análisis comparativo teórico que se tratará de corroborar posteriormente en la evaluación práctica.
- Las diversas técnicas seleccionadas se evaluarán sobre el conjunto de datos propuesto, así como frente a un conjunto de datos de referencia del estado del arte.
- Los resultados se analizarán en profundidad, sacando conclusiones que se consideren de utilidad para aquellos investigadores que quieran hacer uso de este estudio en el futuro.
- Estudio del estado del arte, comparativa teórica de algoritmos de descripción de puntos de interés y selección de los algoritmos más destacados para su evaluación con el *framework* definido sobre los *datasets* propuestos.

Una vez completados todos estos objetivos, se espera que se disponga de un extenso análisis comparativo, así como de un exhaustivo y actualizado marco común de evaluación para las futuras técnicas del estado del arte.

1.3. Estructura de la memoria.

La memoria del proyecto se divide en los siguientes capítulos:

- Capítulo 1. Introducción: introducción, motivación y objetivos del proyecto.

- Capítulo 2. Estado del arte: Introducción e historia de los puntos de interés y estudios previos.
- Capítulo 3 Marco de evaluación: Nuevo conjunto de datos aportado, metodología y selección de métricas de evaluación
- Capítulo 4 Estado del arte de detectores, comparativa teórica y evaluación de algoritmos sobre el marco de evaluación.
- Capítulo 5Estado del arte de descriptores, comparativa teórica y evaluación de algoritmos sobre el marco de evaluación.
- Capítulo 6. Conclusiones y trabajo futuro.
- Referencias y anexos.

Capítulo 2

Estado del arte.

Hoy en día se generan grandes volúmenes de información en forma de imágenes y vídeos. Los segmentos de información que conforman una imagen carecen de identidad semántica si se presentan individualmente. En cambio, cuando son presentados en conjunto, lo que representan estos píxeles puede ser interpretado de manera subjetiva. Este salto semántico supone un nivel de abstracción de gran complejidad y por ellos se trata de una tarea que desde un punto de vista general se encuentra irresuelta, y que en el estado del arte aún se lleva a cabo de forma supervisada.

Actualmente en el área del procesado de imagen se han alcanzado unos resultados relativamente satisfactorios para gran variedad de aplicaciones y tareas complejas [1][8]. No obstante, el desarrollo y funcionamiento de estas aplicaciones, consideradas de alto nivel por estar más próximas al nivel semántico, depende en gran medida del funcionamiento de los algoritmos de más bajo nivel, los más próximos a trabajar a nivel de píxel.

El objetivo de este proyecto de llevar a cabo una evaluación de aquellos algoritmos cuyo objetivo es el de tratar la información contenida en una imagen basándose en puntos, regiones o agrupaciones de interés es de gran relevancia para el estado del arte de cara al uso del mismo como herramienta de selección de técnicas de bajo nivel para sus tareas de alto nivel.

Estas técnicas, a nivel general, buscan un enfoque que le permita alcanzar una descripción de la propia imagen centrándose en puntos o regiones que aportan mayor cantidad de información que otros puntos de su entorno, en lugar de tratar la imagen al completo. Esto se justifica, por lo general, según los principios de la entropía e información, en que las variaciones en un entorno estable presentan gran cantidad de información. Los puntos de interés en una imagen, típicamente, están localizados

[1] [http://changedetection.net/]

donde se produce un cambio en una propiedad de la imagen o varias simultáneamente.

Este capítulo tratará de enmarcar este trabajo en el estado del arte de las evaluaciones de las técnicas de puntos de interés. Para ello las siguientes secciones presentará la siguiente estructura. En primer lugar se hará una breve introducción a los puntos de interés acompañada de un marco histórico y unos conceptos de terminología (sección 2.1). Posteriormente se tratarán a nivel conceptual las dos etapas en las que se dividen las técnicas de puntos de interés, la detección y la descripción, y se presentarán las propiedades que tratan de lograr (sección 2.2). Seguidamente, se incluye una sección en la que se analizarán los estudios relacionados, desglosando entre estudios de referencia, estudios recientes y estudios de aplicación (sección 2.3). Por último, se cerrará el capítulo con unas breves conclusiones sobre el estado del arte.

2.1. Introducción.

2.1.1. Historia.

El uso de los puntos de interés ha tenido un gran crecimiento en las últimas décadas. El aumento de la capacidad de computación en este tiempo ha ido permitiendo a su vez un crecimiento en sus potenciales aplicaciones.

A pesar de esa limitación inicial en los recursos computacionales, la teoría que hay detrás de los métodos de detección del actual estado del arte se remonta a la década de los 70, cuando se implementan los primeros detectores. Sin embargo, las primeras contribuciones se remontan a 1954 [9] cuando, como señala Tuytelaars [10], Attneave postula que la información de la forma de un objeto se concentra en puntos donde dicha forma cambia su "dirección", es decir, zonas de contorno con alta curvatura.

Las primeras implementaciones de detectores estaban orientadas a trabajar con imágenes y dibujos muy estructurados en lugar de escenas naturales. Es el caso por ejemplo del detector basado en la matriz hessiana propuesto por Beaudet en 1978 [10] que obtenía buenos resultados con dibujos simples sin desorden de fondo.

Los trabajos pioneros sobre detección en escenas naturales fueron hechos por Moravec [11] a finales de los 70. El operador de interés de Moravec es una técnica de detección basada en la variación de intensidad en el vecindario local de un píxel. Una versión mejorada de este detector fue propuesta por Harris y Stephens [12], el detector de esquinas de Harris, en el cual están basados algunos métodos actuales.

Algunos de los métodos más populares propuestos en los 70 y 80 buscaban como objetivo principal la precisión en la localización de los puntos. En años posteriores se comenzó a poner el foco conseguir detecciones estables frente a algunas transformaciones como cambios de escala [13] en el sentido de que los mismos puntos puedan ser

detectados ante estas diferentes condiciones de imagen. Para tratar de lograr esta invariancia se propusieron variaciones multi-escala a estos métodos [14] (Crowley&Parker, 1987) y posteriormente Lindeberg [15][16], quien realizó importantes aportaciones a lo que se conoce como espacio-escala. Sobre ellas se han desarrollado también algunos métodos más recientes.

En la última década algunos de los métodos populares surgidos se han enfocado en proveer detecciones rápidas y con bajo coste computacional, como es el caso de FAST [3], y métodos basados en él como AGAST [4] y los detectores de ORB [17] y BRISK [5].

Paralelamente a la detección de puntos de interés se ha estudiado la descripción de estos puntos para la búsqueda de correspondencias básicas entre imágenes. Esta búsqueda se hace menos costosa computacionalmente con el uso de los puntos de interés ya que no es necesaria una comparación exhaustiva, esto es, píxel a píxel, entre las imágenes. Esta propiedad cobra más importancia a medida que la aplicación de estas técnicas se vuelve más intensiva, por ejemplo, cuando se trabajan en indexación con grandes bases de datos de imágenes.

Desde la década de los 80 se han propuesto muchos métodos de descripción de puntos de interés [18]. Entre los descriptores que se pueden encontrar en la literatura existen diferentes aproximaciones para describir el entorno local del punto de interés: con los píxeles de la propia imagen (evaluando niveles de grises), mediante sus histogramas de color o por sus gradientes de orientación. El descriptor más conocido hoy en día, SIFT (*Scale-Invariant Feature Transform*), fue presentado por Lowe en 2004[2].

Los descriptores actuales han alcanzado un nivel de madurez tal que se consideran una herramienta estándar para su uso en muchas aplicaciones. Algunas aportaciones relativamente más recientes, pero también altamente contrastadas, centran su interés en desarrollar métodos más rápidos como es el caso de SURF (*Speed-Up Robust Features*) [1].

2.1.2. Terminología.

Como se ha mencionado, los puntos de interés en una imagen, se refieren a unos puntos con unas características específicas con respecto a su entorno. Debido a ésta información relevante respecto del entorno, para referirse a los puntos de interés, en la literatura es común también el uso de términos como regiones o agrupaciones de interés y características locales (*local features*). Las características locales que, más comúnmente, buscan los algoritmos de detección son esquinas (*corners*), bordes (*edges*) y gotas (*blobs*).

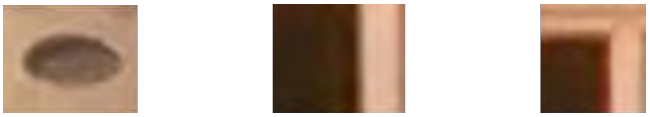


Figura 2.1: Detalle de conceptos de: a) blobs, b) bordes de único gradiente dominante, c) bordes de múltiple gradiente dominante

Los blobs son regiones de la imagen que contienen puntos o píxeles con propiedades similares que los diferencian de su entorno cercano (ver Figura 2.1a)). Los bordes y esquinas son características locales con la propiedad de tener un gradiente predominante en una dirección en el primer caso (ver Figura 2.1 b)) o en varias simultáneamente en el segundo (ver Figura 2.1 c)). Lo que supone que se visualicen como cambios rápidos en color o intensidad.

2.2. Detección y descripción de puntos de interés.

Tal y como se ha comentado anteriormente, las tareas de detección y descripción de puntos de interés suelen ir ligada, aunque en ocasiones detectores y descriptores se presentan en el estado del arte de forma independiente.

En un procedimiento habitual, en primer lugar hay que localizar los puntos de interés en la imagen. La etapa de detección suele consistir en someter a la imagen a determinadas operaciones matemáticas tras las que los puntos de interés resaltan significativamente.

En función de la aplicación o las propiedades buscadas, puede ser más interesante utilizar una técnica u otra. No obstante, independientemente del método utilizado, estos puntos tienen en común que se pueden encontrar de manera sencilla, tienen una posición bien definida dentro de la imagen y su entorno cercano aporta gran cantidad de información local relevante. Todas las detecciones buscan ser capaces de identificar un mismo elemento de interés presente en distintos entornos, puntos de vista y condiciones de imagen, esto es, ser repetibles.

Una vez localizados, el siguiente paso es la descripción de cada punto. Para ello, las aproximaciones se suele usar información del entorno, es decir, local. Cuanta más información del entorno se incluya en la descripción, se podrá conseguir una descripción más discriminativa pero, por contra, con menor repetibilidad [10]. El objetivo es que los distintos elementos detectados en una imagen queden caracterizados de una manera estable, con capacidad distintiva y, sobre todo, repetible. Como ya se ha comentado, el método escogido para la descripción es independiente de qué método

se haya empleado para la detección de los puntos.

Como se ha ido viendo, estos métodos tratan alcanzar los objetivos de repetibilidad y capacidad discriminativa. Para abordar estos objetivos, los métodos presentan una serie de propiedades como lo es la localidad, descrita anteriormente. Analizando por ejemplo el aspecto de la repetibilidad, se busca que dadas dos imágenes del mismo objeto o escena, tomadas bajo distintas condiciones de captura, sea posible detectar e identificar elementos de interés de ambas imágenes con la mayor precisión posible. Esta tarea suele abordarse de dos diferentes maneras: ya sea por invariancia o por robustez.

Cuando son de esperar grandes deformaciones en la imagen, es preferible modelarlas matemáticamente y luego desarrollar métodos para la detección de elementos que sean invariantes a estas deformaciones, esto es, que no se vean afectados por ellas. Por otro lado, para deformaciones relativamente pequeñas, a menudo es suficiente aplicar métodos de detección menos sensitivos, más robustos, a tales deformaciones, lo que puede hacer disminuir la capacidad distintiva de las detecciones.

La distintividad por su parte pretende que los elementos detectados sean lo suficientemente discriminativos como para que sean asociados únicamente a sus elementos correspondientes en otra imagen. Para ello, se proponen la inclusión de distintas características del entorno en las descripciones, variaciones en los tamaños y formas del entorno utilizado en la descripción, o diversas formas de presentar las descripciones.

Propiedades comunes a la repetibilidad y a la distintividad son la cantidad de elementos detectados o precisión en su localización. Propiedades que a su vez se ven limitadas cuando se trata de aplicaciones en tiempo real por la eficiencia de los algoritmos.

Estas propiedades deseables para los métodos de detección de puntos de interés que hemos definido (repetibilidad, distintividad, localidad, cantidad, precisión y eficiencia) fueron primeramente descritas por Tuytelaars y Mikolajczyk en uno de los estudios de referencia en este área [10].

2.3. Estudios.

2.3.1. Estudios de referencia.

Entre las evaluaciones de referencia de detectores y descriptores de puntos de interés que se pueden encontrar en la literatura, hay que nombrar en primer lugar los trabajos publicados entre 2003 y 2004 por Mikolajczyk y Schmid [19][18].

En [19] realizan una evaluación de cinco detectores de puntos de interés invariantes a escala y a otras transformaciones. Los detectores incluidos en la evaluación fueron

Harris-Laplace [19][20], Hessian-Laplace [19][20], MSER [21], DoG [22] y Regiones salientes, para lo cual midieron el rendimiento de estos detectores ante cambios en las condiciones de la imagen tales como cambio de escala, cambio de punto de vista, emborronado (*blur*), rotación, cambio de iluminación y compresión JPEG. En muchos casos MSER obtenía las mejores puntuaciones seguido por Hessian-Affine y Harris-Afine.

En [18], el estudio homólogo de descriptores, propusieron además una colección de imágenes para realizar la evaluación, con las diferentes transformaciones geométricas y fotométricas mencionadas anteriormente, que además, posteriormente se ha consolidado como un *dataset* de referencia en el estado del arte para *test* y evaluaciones de nuevos detectores y descriptores. La evaluación experimental contenía diez descriptores, SIFT [2], GLOH[18], Shape Context [23], PCA-SIFT [24], Spin Images [25], Steerable filters [26], differential invariants[27] , Complex Filters [28], Moment Invariants[29], y Correlación Cruzada CC [18]. En este caso GLOH y SIFT son los que obtuvieron el mejor rendimiento.

2.3.2. Estudios recientes.

Las evaluaciones anteriores, como se ha mencionado, se han establecido como referencia en el estado del arte, por lo que es posible encontrar algunas evaluaciones similares cuando son presentados nuevos detectores y descriptores, y que estos sean comparados con métodos populares y reputados como SIFT y SURF [30][10].

Una tercera evaluación realizada en 2012 por Micolajczyk y Miksik [31], conteniendo a la vez detectores y descriptores, centra el interés en este caso en la búsqueda rápida de correspondencias entre imágenes (*fast feature matching*). La evaluación contempla los métodos ORB [17] y BRISK [5], que constan de detector y descriptor, y los descriptores BRIEF [32], SIFT [2], SURF [1], MROGH [33], MRRID [33] y LIOP [34]. Para todos estos descriptores la evaluación se realiza con puntos de interés obtenidos mediante el detector de SURF para conseguir resultados de descripción comparables. Además, el estudio se completa con una breve comparación de tiempos de ejecución de detectores de referencia del estado del arte como FAST [3], SIFT, SURF, DoG, etc. El resultado en este caso fue, como apuntan otros estudios como [17], que FAST y los métodos basados en éste detector son uno o dos órdenes de magnitud más rápidos que detectores como los de SURF y SIFT.

Un trabajo muy similar a este último fue presentado también en 2012 por Heinly y Dunn [35] en el que proponen como añadido una colección propia de imágenes para realizar la evaluación.

2.3.3. Estudios orientados a aplicaciones.

Es posible encontrar también estudios orientados a analizar el funcionamiento de estos métodos para tareas más específicas, como por ejemplo, aplicaciones de fotogrametría [36], localización y mapeado simultáneos en robótica (SLAM) [37], reconstrucción 3D [38][39]. Aunque no se va a centrar el foco del trabajo en ellos, se propondrá como trabajo futuro establecer una relación directa entre las propiedades evaluadas sobre los puntos de interés y su relación con un mejor funcionamiento en aplicaciones concretas.

2.4. Conclusiones.

Como se ha podido ver, en la última década se ha desarrollado una variedad de nuevas técnicas de detección y descripción de puntos de interés. Mientras tanto los trabajos de evaluación, como los conjuntos de datos empleados en ellos, no se han ido renovando al mismo ritmo. Es una práctica habitual que los autores de los nuevos algoritmos efectúen una evaluación en la que se compara el nuevo método con los métodos mas contrastados y punteros, pero esto se suele realizar en base a los mismos conjuntos de datos y los trabajos de evaluación que se han presentado en este capítulo. Estos hechos son los que motivan el desarrollo de este proyecto que pretende actualizar y complementar los conjuntos de datos existentes, así como ofrecer un marco de evaluación para las nuevas técnicas.

Capítulo 3

Marco de evaluación.

El principal objetivo de esta sección es fijar un marco común de evaluación de detectores y descriptores de puntos de interés que actualice, de homogeneidad y mejore a los anteriormente propuestos.

El primer paso es proponer un conjunto de datos (*dataset*) de imágenes para la llevar a cabo evaluación comparativa de los algoritmos de detección y descripción. Para ello, el *dataset* debe reproducir transformaciones de imagen típicas que a priori puedan suponer dificultades para el funcionamiento de los algoritmos. Como se ha mencionado, los trabajos publicados K. Micolajczyk y T. Tuytelaars en 2004 [19][18] son un punto de referencia en el estado del arte, así como el dataset de evaluación que presentaron. Una de las motivaciones del nuevo conjunto de imágenes que se propone es consecuencia del gran uso que se le ha dado a ciertas secuencias del *dataset* de referencia, y que ha hecho que comience a haber sobre ajuste de la técnicas propuestas al propio *dataset*. Esta saturación, puede resultar en técnicas con grandes puntuaciones en evaluaciones sobre el mencionado *dataset* que no se corroboran con el comportamiento de las mismas en situaciones reales. Sin embargo estos problemas no quitan validez al enfoque original, por lo que la grabación del nuevo conjunto de imágenes estará inspirada en ideas similares.

En este tiempo, la calidad de la fotografía digital ha mejorado notablemente, por lo que otra de las motivaciones surge de la necesidad de que el marco de evaluación permita trabajar con todos los rangos de calidades disponibles hoy en día. Esta mejora se ve reflejada, no solo en la resolución de las imágenes (que se multiplica hasta por diez), sino que también se reduce la aparición de elementos indeseables como artefactos y ruido que podrían enmascarar resultados. Partir de imágenes de alta calidad supone además tener siempre la posibilidad de reducirlas o recortarlas para manejar volúmenes menores de información si fuese necesario.

El segundo paso es establecer un marco de evaluación común en el aspecto métrico, lo cual está ligado a fijar una metodología y proveer de los criterios de evaluación adecuados en cada caso. De cara a este objetivo se analizaran y seleccionarán las métricas de evaluación más recomendadas.

Este capítulo se organizará como sigue. En primer lugar se presentará el mencionado dataset (sección 3.1), donde se desglosarán los criterios y objetivos seguidos para el diseño del conjunto de datos, el proceso llevado a cabo para la generación y un resumen del conjunto resultante. Posteriormente se desarrollará la metodología de evaluación propuesta (sección 3.2), donde se desglosarán las métricas escogidas tanto para la evaluación de las técnicas de detección como de descripción.

3.1. Dataset.

3.1.1. Objetivos.

Tanto los algoritmos de detección como los de descripción tratan de proveer puntos de interés robustos frente a una serie de transformaciones típicas de imagen. El dataset constará de secuencias de imágenes que presentarán estas condiciones en distintos niveles de afectación. La selección de las propiedades de estas secuencias se realizará en base a distintos criterios.

A partir del estudio del estado del arte de detectores y descriptores se puede realizar una primera selección de transformaciones de imagen frente a las que los diversos algoritmos pretenden conseguir invariancia mediante la implementación de distintas técnicas. Las tres más resaltables en este sentido son cambios de escala, rotación y transformaciones afínes (cambios de punto de vista).

En segundo lugar, los conjuntos de datos más populares como el de Micolajczyk[1], presentan secuencias con propiedades que también son objetivo de este marco de evaluación como cambios de iluminación, y borrosidad en la imagen (*blur*) así como secuencias que combinan cambios de escala y rotación.

Como se ha comentado, con el objetivo de renovar y complementar los *datasets* existentes, y como continuación del punto anterior, adicionalmente se crearán secuencias en las que las propiedades anteriores no se presentan individualmente sino combinadas, como por ejemplo *blur* y cambio de iluminación, *blur* y cambio de punto de vista, o cambio de iluminación combinado con cambio de punto de vista.

Por último, para enriquecer el conjunto de datos desde un punto de vista de aplicación, se propondrán secuencias en las que estas variaciones se produzcan solo sobre

[1] `http://www.robots.ox.ac.uk/~vgg/research/affine/`

un elemento de interés de la escena, modelando situaciones tales como la aparición sombras y *blur* sobre ciertas partes o elementos de la imagen. Esto permitirá evaluar puntos de interés y sacar conclusiones sobre su funcionamiento en aplicaciones que pueden sufrir este tipo de situaciones, como pueden ser aplicaciones de seguimiento de objetivos.

El resultado final serán por tanto 12 secuencias con las siguientes propiedades:

Transformación única	Global	Cambio de iluminación
		Cambio de punto de vista
		Blur
	Objetivo	Blur sobre un elemento de la imagen
		Cambio de iluminación sobre un elemento de la imagen
		Ensombrecido con expansión en área
Combinación de transformaciones	Global	Blur con cambio de iluminación
		Escala con rotación
		Cambio de punto de vista con cambio de iluminación
		Cambio de punto de vista con Blur
	Objetivo	Cambio de iluminación y blur sobre un elemento de la imagen
		Ensombrecido con expansión en área y blur

Tabla 3.1: Características desglosadas del dataset generado. Propiedades: iluminación, *blur*, punto de vista, escala y rotación. Elemento de afectación: toda la escena (global) o solo el objeto de interés (objetivo).

3.1.2. Proceso

Todas las capturas de imágenes necesarias para la creación del *dataset* se han realizado con una cámara de 7.99 Megapíxeles con enfoque automático y han sido posteriormente editadas cuando ha sido necesario mediante la herramienta de prototipado Matlab.

A continuación se detallan los procesos que han sido necesarios para conseguir cada tipo de propiedad. Para lo cual se explicarán métodos de obtención de las secuencias más características en cada caso a modo de ejemplo. Como son: cambio de punto de vista, cambio de escala y rotación, blur y cambio de iluminación. También se expondrán las técnicas para extraer máscaras de los objetos de interés y las estrategias llevadas a cabo para la combinación de propiedades.

3.1.2.1. Cambio de punto de vista

La captura de las seis fotos de la secuencia se realizó desde una distancia de aproximadamente 1,5 metros hasta el plano objetivo, moviendo la cámara en pasos de 15 grados, por lo que el cambio de punto de vista máximo, que se produce entre la primera y la sexta imagen de la secuencia, es de 75 grados.

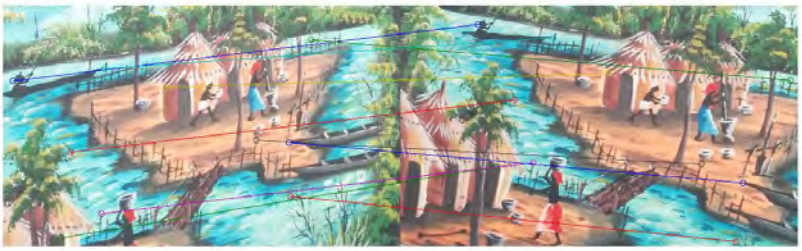


Figura 3.1: Anotación de correspondencias para cambio de punto de vista

Las secuencias con este tipo de transformaciones no requieren una edición posterior, pero sí un proceso de preparación de homografías que permitan modelar el movimiento de la cámara de una imagen respecto a otra. Con este modelado se obtiene los que se conoce como *groundtruth,* y que será la información que permita posteriormente, junto con las métricas, estimar la corrección o no de los resultados de cada algoritmo.

Para realizar el cálculo de homografías se adaptó una herramienta de Matlab que permitía llevar a cabo el proceso de forma precisa y supervisada. La técnica escogida involucra un primer proceso de anotación. A partir de una herramienta en Matlab se anotan manualmente correspondencias de puntos desde la primera foto de la secuencia a todas las demás, como se puede ver en la Figura 3.1. Esto hace que la precisión con la que se obtengan esas correspondencias este directamente relacionada con la calidad de las posteriores homografías. Por ello, se ha realizado una adaptación de dicha herramienta para conseguir una ampliación de la imagen que permita escoger los puntos a nivel de píxel.

A partir de los datos de anotación, las homografías se han calculado haciendo uso del algoritmo RANSAC [40] para evitar que aquellas anotaciones erróneas distorsionasen el resultado. El resultado final es el citado *groundtruth* necesario para la evaluación, ver Figura 3.2.

Figura 3.2: Cálculo del *groundtruth* a) Imagen de referencia, la primera imagen de la secuencia b) Imagen sobre la que se desea calcular la transformación, c) Ejemplo de la aplicación del *groundtruth* sobre la imagen de referencia.

3.1.2.2. Cambio de escala y rotación

La secuencia de escala con rotación consiste en un objeto que rota y cambia de escala sobre un fondo estático, esto es, la transformación solo se aplica al elemento de interés. Las capturas del objeto se han realizado desde una distancia entre la cámara y objeto de 70 centímetros (la más alejada) hasta 20 centímetros (la más cercana) en pasos de 10 centímetros. A su vez, en cada paso, el objeto se somete a una rotación de 45 grados. Las imágenes originales del objeto se han capturado sobre un croma para poder extraer una máscara del mismo, ver Figura 3.3. Esto permite generar el fondo estático y realizar la evaluación únicamente en las partes de la imagen que están sometidas a los efectos de escala y rotación.

Figura 3.3: a) Objeto sobre croma, b) Máscara del objeto, c) Imagen final

Para la evaluación también es necesario el cálculo de las homografías que modelan las transformaciones mediante la herramienta de anotación y el algoritmo de cálculo de homografías a partir de esos datos, el cual ha sido presentado en la sección anterior, ver Figura 3.4. Destacar que en este caso la imagen de transformación presenta ese aspecto porque al pasar de una imagen de menor escala a una de mayor, parte de la información es nueva y no está definida.

Figura 3.4: Proceso de cálculo del *groundtruth*: a) Anotación de correspondencias entre imágenes, b) Resultado del proceso: la imagen objetivo se modela a partir de la imagen de referencia.

3.1.2.3. Blur: Suavizado gaussiano progresivo

La generación de las secuencias con efectos de *blurring* se ha realizado aplicando suavizados gaussianos, de media 0 y $\sigma = 1$, en los que el factor σ del núcleo gaussiano aumenta progresivamente. De tal manera que en las secuencias, de seis imágenes cada una, la primera imagen es la original capturada por la cámara y en las posteriores se simulan cinco niveles de afectación definidos como $\{\sigma, 2\sigma, 3\sigma, 4\sigma, 5\sigma\}$. El efecto se puede observar en el detalle mostrado en la Figura 3.5. En el caso de las secuencias en las que únicamente es un elemento de la imagen el que se ve afectado, el suavizado se aplica mediante una máscara, sólo sobre ese elemento, haciendo uso de nuevo del croma y de las máscaras de segmentación, igual que en la sección 3.3.

Figura 3.5: Detalle secuencia con suavizado gaussiano: a) Imagen original, b) Imagen mayor grado de afectación

Se ha modelado además un caso especial de este tipo de transformaciones en el que un elemento de la imagen se ve afectado por un emborronado lineal producido por el movimiento del elemento. Para ello se ha utilizado un filtro predefinido en Matlab que simula borrosidad por movimientos de cámara.

Figura 3.6: Blur por movimiento lineal: a) Imagen original, b) Objeto emborronado.

3.1.2.4. Cambio de iluminación

Para modelar un cambio de iluminación mediante Matlab se ha aplicado una transformación lineal sobre las componentes de color de las imágenes del tipo: $I_n = \alpha I_0 - \beta$, donde α toma valores desde 1 en la imagen original hasta 0,4 en la más oscurecida y β desde cero (imagen original) hasta 10 para la última imagen de cada secuencia.

De nuevo, cuando es sólo un elemento el que se ve afectado por los cambios de iluminación, el montaje se ha hecho gracias a la captura mediante croma y el cálculo posterior de una máscara. Un ejemplo de la secuencia generada se muestra en la Figura 3.7.

Adicionalmente, en una de las secuencias no ha sido necesario modelar el cambio de iluminación ya que las imágenes se han capturado en esas condiciones reales. La secuencia parte de un estado de iluminación completo en la que paso a paso se va ensombreciendo más área de la imagen.

Figura 3.7: Resultado del cambio de iluminación: a) Imagen original, b) Imagen más oscura de la secuencia

3.1.2.5. Combinación de propiedades

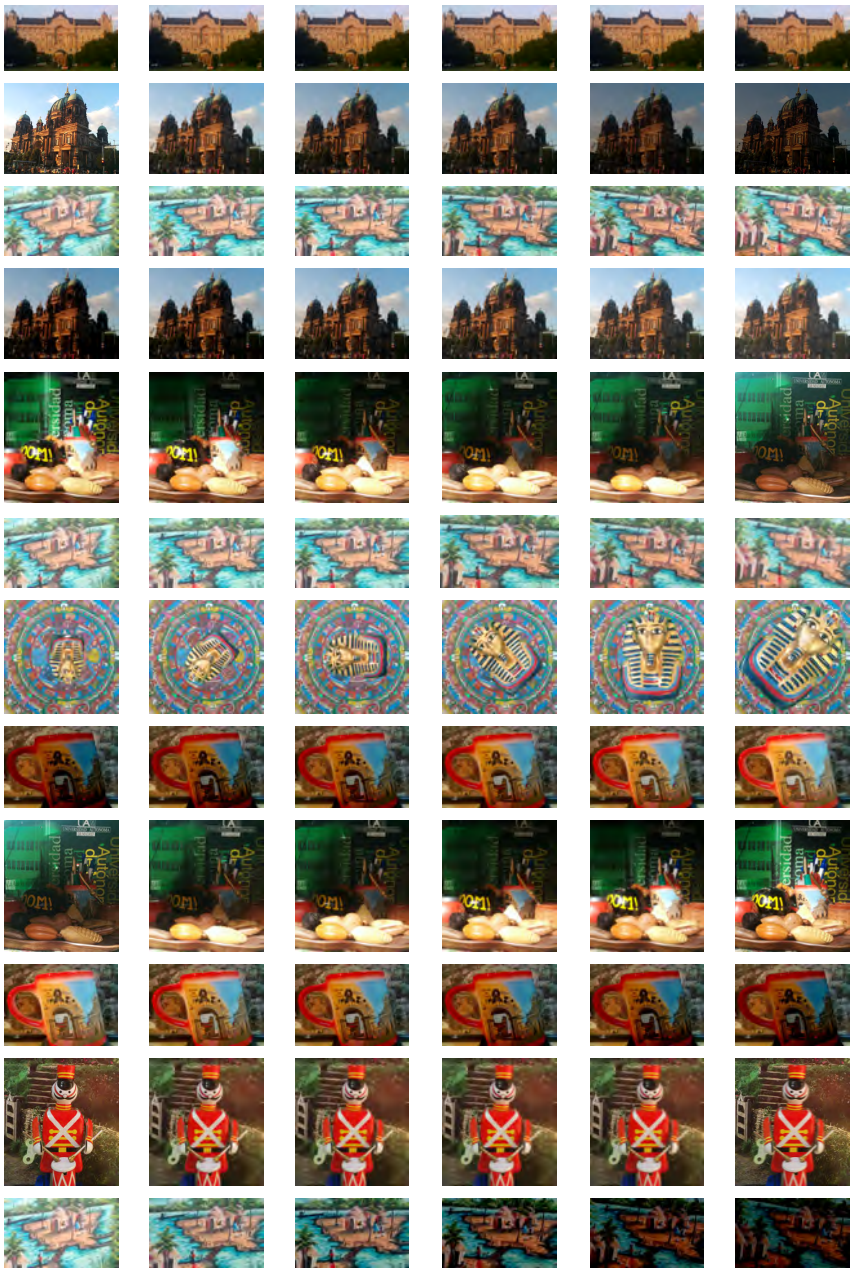
Los procesos para obtener las secuencias que se ven afectadas por una combinación de propiedades son una combinación de los métodos ya explicados para cada propiedad.

Figura 3.8: Cambio de iluminación y blur: a) Imagen original, b)Imagen clara y blur

En el caso de las secuencias que combinan cambio de iluminación con blur, éstas,

parten de la imagen más oscura sin blur, hasta la imagen más aclarada pero con un blur máximo. El motivo de realizarlo de esta manera es el de simular un efecto típico de cámara que ocurre cuando se produce un aumento de iluminación, por el que se causa un desenfoque de la imagen en mayor o menor medida. Un ejemplo de este tipo de transformaciones se puede ver en la figura 3.8.

3.1.2.6. Resultado final



3.1.3. Comparativa con el dataset de referencia

Adicionalmente, se realizará la evaluación de los algoritmos sobre uno de los conjuntos de datos más populares, publicado por K. Micolajczyk[2], en el que están inspiradas algunas características del dataset aportado para este proyecto.

El dataset consta de cinco cambios diferentes en las condiciones de imagen: blur, cambio de iluminación, compresión JPEG, cambio de punto de vista y zoom combinado con rotación. En los casos de blur, cambio de punto de vista y rotación combinado con zoom, el dataset aporta dos diferentes tipos de escena para cada transformación. Para cada una de ellas, en una de las escenas predominan las regiones homogéneas con bordes bien definidos y la otra escena consta de regiones heterogéneas y texturas repetitivas.

Cada una delas secuencias consta de seis imágenes de aproximadamente 800x640 píxeles sobre las que se producen transformaciones fotométricas y geométricas.

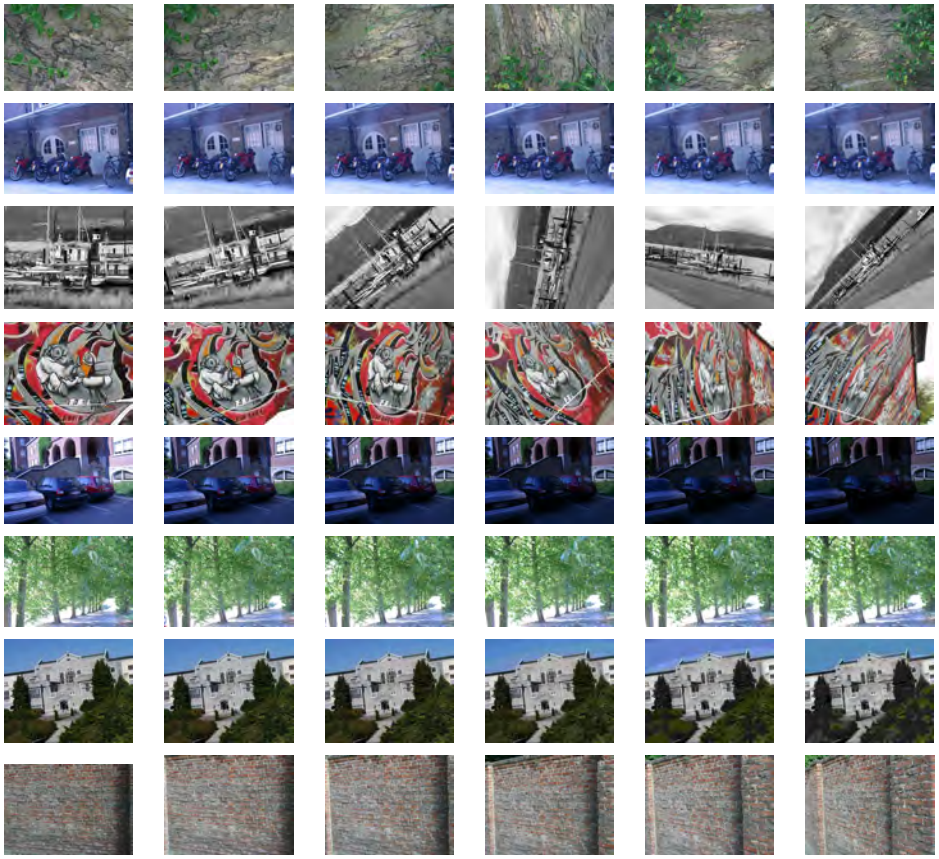
Para el test de cambio de punto de vista, las dos secuencias parten de una imagen plana capturada desde una posición frontal y en las sucesivas imágenes se va variando el punto de vista hasta una posición lateral con 60 grados de diferencia con la original.

Los cambios de escala y blur, según el autor, están obtenidos variando el enfoque y zom de la cámara. De manera similar, el cambio de iluminación está obtenido modificando la apertura de la cámara. Por último la secuencia de compresión JPEG está obtenida variando la calidad de la imagen desde un 40 % hasta un 2 % mediante una herramienta de edición de imágenes.

Salvo la secuencia de compresión JPEG, todas las demás están acompañadas de homografías debido a los movimientos de cámara. Las homografías están obtenidas según los autores mediante un proceso de anotación manual de correspondencias y un posterior paso de procesado, mediante un algoritmo de estimación de homografías. El resultado se muestra en la sección 3.1.3.1.

[2] http://www.robots.ox.ac.uk/~vgg/research/affine/

3.1.3.1. Dataset de referencia



3.2. Métricas de evaluación

El objetivo de esta sección es establecer un marco de evaluación común de detectores y descriptores en el aspecto métrico, para lo que se seleccionarán las medidas que se consideren más adecuadas para cada caso y se desarrollará una metodología de evaluación en base a esas métricas. A continuación se presentarán divididas en métricas para la evaluación de detectores y métricas para la evaluación de descriptores.

Previamente es necesario conocer el dataset descrito anteriormente. El conjunto de datos se estructura en categorías, o propiedades, a evaluar. Dentro de cada categoría, una imagen se fija como la de referencia, y el resto forman el conjunto a evaluar. La información de *groundtruth* generada permite conocer la ubicación exacta de cada punto de la imagen de referencia en cualquiera de las imágenes de la categoría.

3.2.1. Evaluación de detectores

3.2.1.1. Metodología detección

La intención de establecer un marco de evaluación referente y común, obliga a fijar una metodología determinada para llevar a cabo la evaluación. Dicha metodología se estructura en en una serie de pasos.

1. Se lleva a cabo la detección de puntos de interés para un par de imágenes. En este caso, una de ellas será siempre la imagen de referencia, mientras que la otra será una de las restantes de la categoría.

2. Los puntos detectados en la imagen de referencia se trasladan, haciendo uso de la información de *groundtruth*, a sus correspondientes posiciones en la imagen con la que se está evaluando.

3. Haciendo uso de la información de localización de las detecciones se obtienen una serie de valores que se definen como sigue:

   True Positives (TP): Se obtiene un *true positive* cuando la detección transformada desde la imagen de referencia coincide, espacialmente, con una detección en la imagen evaluada.

   False Negatives (FN): Se definen como *false negative* a todos los puntos detectados en la imagen de referencia que no obtienen ninguna correspondencia en la imagen evaluada.

4. Haciendo uso de estos datos, se aplica la métrica fijada y se obtienen valores para cada una de las imágenes de cada una de las categorías.

3.2.1.2. Métrica detección

Aunque algunos detectores proveen puntos de interés con otros parámetros como pueden ser la escala característica o su orientación dominante, el parámetro que se tendrá en cuenta para la evaluación no es otro que la posición de los puntos.

En un caso ideal, el algoritmo de detección será capaz de localizar una serie de puntos en la imagen, y mantener esas detecciones aunque la imagen se vea afectada por los efectos descritos de cambio de iluminación, punto de vista, escala, etc. Sin embargo, el efecto esperado es que a medida que el nivel de afectación de la imagen sea más severo, por ejemplo, los cambios de escala sean más notables, estas detecciones se vayan perdiendo. Este efecto es el que se desea medir en la evaluación. Para ello, hay que seleccionar una técnica que permita medir en que grado, los puntos detectados en

la imagen original o de referencia (no afectada), están siendo correctamente detectados de nuevo en las imágenes afectadas.

En la subsección anterior se han definido los valores de TP y FN. Ambos se han obtenido en base a la información de localización de los puntos en la imagen a evaluar. Recordar que los puntos de la imagen de referencia son transformados a sus posiciones en la imagen a evaluar mediante la información de *groundtruth*. Sean dos puntos cualesquiera, uno de la imagen de referencia transformado p, y otro de la evaluada q, se podía asignar al punto p una etiqueta de clasificación l, en base a las definiciones anteriores como TP o FN según 3.1.

$$l_{(p,q)} = \begin{cases} TP & d(p,q) < th \\ FN & d(p,q) \geq th \end{cases} \tag{3.1}$$

Siendo $d(p,q)$ las distancia euclídea entre las coordenadas de ambos pares de puntos, y th la distancia fijada para considerar que dos puntos se han detectado en la misma posición. Pese a que en el estado del arte se hacen barrido del parámetro th para ver su efecto, en esta evaluación se ha definido a un valor fijo de 3. Éste valor permite considerar el mismo punto a todas aquellas detecciones que se encuentren en contacto con conectividad-8 .

En base al recuento final de etiquetas asignadas a los distintos puntos, se pasa a definir la métrica. Habida cuenta que lo que se pretende medir es la robustez de las detecciones, y no algún tipo de precisión ya que no se realiza ninguna asociación por parte del algoritmo, se ha considerado que la métrica que mejor refleja el comportamiento de un algoritmo en este sentido es la definida como *RECALL* (3.2). En función del valor de esta métrica a lo largo de una categoría, se podrá determinar que un algoritmo es más robusto o menos antes la propiedad que define la categoría.

$$RECALL = 100x\frac{\#TP}{\#TP + \#FN} \tag{3.2}$$

3.2.2. Evaluación de descriptores

3.2.2.1. Metodología descripción

Al igual que se ha procedido para las técnicas de detección, se va a fijar una metodología de evaluación para proceder a evaluar las distintas técnicas de descripción del estado del arte. Para ello será necesario definir dos premisas que se han considerado para realizar la evaluación.

- Solo se pretende evaluar la capacidad del descriptor sin información previa de

la imagen salvo la ubicación espacial del punto de interés.

- Se trata de evitar que las técnicas de detección ocluyan los resultados de la descripción por lo que se define un conjunto de puntos detectados que serán los usados por todas las técnicas de descripción.
- La técnica de asociación (matching) entre descriptores será la definida en cada caso por el algoritmo.

En base a estas premisas, se ha diseñado una metodología que permite evaluar las técnicas de descripción.

1. Se parte de un conjunto de detecciones comunes. Dichas detecciones se podrían ir perdiendo en las distintas imágenes de la categoría e impedirían discernir errores de detección frente a errores de descripción. Por ello, el conjunto se ha definido como sigue:

   a) Se ha realizado la detección de un set de puntos en la imagen de referencia, $\{p_k\}$ con $k = 1...K$, siendo K el número total de detecciones de la imagen.

   b) Dicho set de puntos se ha transformado, haciendo uso de la información de *groundtruth*, a cada una de las imágenes de la misma categoría, generando nuevos sets de puntos, $\{q_n\}$ con $n = 1...N$. Con esto se garantiza que se conservan todas las detecciones siempre que sean visibles ($RECALL$ de detección = 1).

   c) El número de detecciones trasladadas define el parámetro $\#correspondencias = N$, que se utilizará posteriormente.

2. Se obtienen los vectores de descripción a partir de cada set de detecciones y cada imagen, generando los correspondientes sets de vectores de descripción $\{\overrightarrow{d_k^p}\}$ y $\{\overrightarrow{d_n^q}\}$.

3. Aplicando las métricas de asociación definidas para cada algoritmo, se obtiene un conjunto de asociaciones entre descriptores $M = \{\overrightarrow{d_i^p}, \overrightarrow{d_j^p}\}$ cuyo cardinal define el parámetro $\#asociaciones = |M|$.

4. Cada asociación de descriptores define una correspondencia entre vectores de descripción, y por ende, de los puntos asociados a dichas descripciones. De igual manera, la información de *groundtruth* define como se asocian los puntos en la imagen de referencia con los de cada imagen de las de su categoría. En base a eso, se definen dos conceptos.

True Poistives (TP): Al igual que para la detección, se definen como *true positives* a aquellos puntos de la imagen de referencia, que al ser transformados según la información de groundtruth coninciden con un punto en la imagen a evaluar. Sin embargo, en esta ocasión solo se considerará positivo cuando la coincidencia se obtenga entre puntos cuyos descriptores se habían asociado en la etapa anterior.

False Positives (FP): En esta ocasión si es posible definir *false positives* ya que se realizan asociaciones entre puntos por parte del algoritmo. Dichas asociaciones podrían definirse entre puntos que según el groundtruth pertenecen a la misma localización, lo que originaría los anteriormente definidos true positives. Por otra parte, si las asociaciones se realizan entre dos puntos que al transformar las coordenadas del obtenido en la imagen de referencia, se obtienen unas coordenadas distintas a las del punto con el que se había asociado, se considera esa asociación como un false positive.

5. En base a todos estos valores obtenidos, se fijarán una serie de métricas que permitirán evaluar el comportamiento de las distintas técnicas.

3.2.2.2. Métricas descripción

Los descriptores, en conjunto con sus técnicas de asociación, definen correspondencias entre pares de detecciones en distintas imágenes. En función de las posiciones de esas detecciones y de la información de *groundtruth*, será posible definir según la metodología antes expuesta, una serie de métricas que evalúen cuantitativamente los resultados de estos algoritmos.

En un caso ideal, siempre que dos detecciones se hayan realizado en el mismo punto de una imagen, hecho al que se ha obligado en la definición del conjunto de detecciones, los descriptores deberían ser capaces de establecer una correspondencia entre ellos de manera única e inequívoca. Sin embargo, a medida que las imágenes se van viendo afectadas por las propiedades a evaluar, las descripciones se vuelven menos discriminativas y comienzan a aparecer asociaciones erróneas, o simplemente dejan de realizarse las asociaciones.

Anteriormente se han definido una serie de valores, como son TP y FP. Estos valores se estiman sobre cada par de descriptores $(\overrightarrow{d_i^p}, \overrightarrow{d_j^p})$, que llevan asociados unos puntos (p_i, p_j) con unas localizaciones espaciales determinadas. Transformando las coordenadas de las detecciones de la imagen de referencia, cada para de descriptores asociados, y por lo tanto cada punto de la imagen de referencia (y de la evaluada) se puede etiquetar, l, según se muestra en 3.3.

$$l_{(\overrightarrow{d_i^p},\overrightarrow{d_j^p})} = \begin{cases} TP & d(p_i, p_j) < th \\ FP & d(p_i, p_j) \geq th \end{cases} \tag{3.3}$$

Siendo de nuevo $d(p_i, q_j)$ las distancia euclídea entre las coordenadas de ambos pares de puntos, y th la distancia fijada (de nuevo a 3) para considerar que dos puntos se han detectado en la misma posición.

En base de nuevo al recuento de etiquetas, se puede definir una primera métrica. Esta métrica permite definir con cuanta precisión se realizan las asociaciones entre puntos, esto es, de entre todas las asociaciones cuántas son correctas y cuantas erróneas. Con este objetivo se va hacer uso de la métrica $1 - Precisión$ que se define según 3.4.

$$1 - Precisión = \frac{\#FP}{\#FP + \#TP} \tag{3.4}$$

Sin embargo, ésta métrica no permite evaluar comparativamente todos los descriptores. Algoritmos que generen pocas correspondencias pero todas ellas correctas, obtendrá una mejor precisión que aquellos que generen muchas pero no todas correctas. Sin embargo, el aspecto de obtener un mayor número de correspondencias, considerado como un aspecto positivo, no se refleja mediante la métrica anterior. Se ha recurrido al estado del arte para ver como enfrentan este problema, y en base a lo presentado por [18], se ha definido TA, tasa de asociaciones, según (3.5).

$$TA = \frac{\#asociaciones}{\#correspondencias} \tag{3.5}$$

Capítulo 4

Detectores.

Según lo expuesto en el Capítulo 2, la etapa de detección suele consistir en someter a la imagen a determinadas operaciones matemáticas tras las que los puntos de interés resaltan significativamente. En el estado del arte se han propuesto numerosas técnicas en los últimos años, y como se ha comentado en el Capítulo 1, resulta complicado discernir entre cuales presentan mejores comportamientos y ante que propiedades.

En este capítulo se va llevar a cabo una categorización de las técnicas del estado del arte (sección 4.1) que permitirá a su vez realizar una pre-selección de los algoritmos más relevantes dentro de cada categoría. Los algoritmos seleccionados serán analizados teóricamente en la misma sección. Dicho estudio teórico se complementará con un análisis comparativo desde el punto de vista teórico (sección 4.2), donde se ha desarrollado una evaluación teórica de cada una de las técnicas escogidas frente a las distintas propiedades de interés de los detectores. Los algoritmos seleccionados serán evaluados en el marco de evaluación propuesto (sección 4.3). El capítulo se cierra con unas conclusiones (sección 4.4) sobre los resultados obtenidos.

4.1. Categorización y selección de técnicas.

En base a un exhaustivo análisis de las técnicas de detección del estado del arte, se va a proponer una categorización de las técnicas del mismo. Dicha categorización facilitará la selección de los algoritmos a evaluar, recurriendo a un balance entre algoritmos más contrastados y algoritmos que venden mejores resultados.

El criterio escogido para la clasificación ha sido la interacción con el medio que realizan las técnicas, previamente a la selección de los puntos de interés. Atendiendo a este criterio, se pueden generar dos grandes categorías en el estado del arte:

- Operadores matemáticos: en estos algoritmos, con el fin de destacar aquellos

puntos que mejores propiedades de repetibilidad pueden presentar, se aplica un operador matemático a la escena completa, o a partes de ella, y en función de la respuesta se toma la decisión de la detección. En este grupo se pueden catalogar detectores tradicionales como Harris (1988) u otros más novedosos como D1 (2013).

- Detectores de entorno: en esta ocasión, las técnicas de esta categoría buscan candidatos a punto de interés en función de su relación con el entorno. En esta categoría se encuadran las propuestas más recientes del estado del arte, como AGAST (2010) u ORB (2011).

Se ha elegido este criterio, ya que garantiza no sesgar el estado del arte. La clasificación ajustada a este criterio permite que la selección de algoritmos esté acotada sin perder riqueza, esto es, garantizando que los métodos más relevantes estén presentes, añadiéndoles aquellos más innovadores. A continuación se va a presentar cada una de las categorías, y dentro de ellas se van a describir brevemente cada uno de los algoritmos escogidos para la posterior evaluación.

Por último, y ya que podría dar lugar a confusión con la categorización realizada, se va a pasar a definir el espacio-escala (E-S). La diferencia entre esta técnica y la aplicación de operadores matemáticos es que esta técnica no origina detecciones, sino espacios sobre los que luego aplicar cualquiera de las dos técnicas antes descritas para realizar las detecciones, por lo que ambas pueden o no incluir la característica de E-S. La técnica del E-S se fundamenta como sigue:

Se pueden conseguir puntos de interés invariantes a escala mediante la búsqueda de los mismos en las múltiples escalas resultantes de la construcción del denominado espacio-escala de la imagen. Partiendo de una imagen $I(u, v)$, esta construcción $L(u; v; \psi)$, 4.1,estará formada por una familia de imágenes suavizadas resultantes de la convolución de la imagen original con un filtro gaussiano en distintas escalas [15].

$$L(u; v; \psi) = g(x, v, \psi) * I(u, v) \tag{4.1}$$

Se utiliza un kernel normalmente gaussiano porque no introduce nuevas estructuras a la imagen, sino que ciertas estructuras prevalecen y otras aumentan su significancia. El filtro gaussiano queda definido en 4.2.

$$g(x, v, \psi) = \frac{1}{2\pi\psi^2} e^{\left(-\frac{u^2+v^2}{2\psi}\right)} \tag{4.2}$$

Conforme aumenta el suavizado, aumenta lo que se llama escala, que se suele establecer como proporcional al filtro gaussiano, en la que para $i = 0$ no existe suavizado,

definiendo así 4.3.

$$\psi_i = i\sigma^2 \tag{4.3}$$

Una vez se dispone del E-S construido, es el momento de aplicar la técnica en cuestión, para la detección final de los puntos de interés en cualquiera de las imágenes del espacio-escala.

A continuación, se va a pasar a describir, por categoría, los métodos que se han considerado más relevantes en el estado del arte.

4.1.1. Operador matemático.

Esta categoría se ha dividido a su vez en una serie de subcategorías. El objetivo aquí, es el de aumentar la granularidad de la división en categorías de tal manera que se garantice que la selección de métodos incluye técnicas lo más variadas posibles. La división se ha realizado en función del tipo de operador matemático que aplican, a saber:

- Matriz de covarianza: Detector de Harris, Multi-scale Harris, Harris-Laplace y Harris-Affine.
- Matriz hessiana: Detector de Hessian, Multi-scale Hessian, Hessian-Laplace, Hessian Affine y SURF.
- Matriz laplaciana: SIFT y D1.

Pese a que las técnicas anteriores pueden incluir algún otro proceso adicional, se han clasificado en función de su operador principal. Según estas subcategorías, se va a pasar a explicar todos ellos a continuación.

4.1.1.1. Métodos de matriz de covarianza.

Harris detector.

El detector de esquinas de Harris, propuesto por Harris y Stephens [12], es precursor, e incluso forma parte, de muchos otros detectores y descriptores de esquinas que han sido usados con buenos resultados en aplicaciones como reconocimiento y categorización de objetos [41][42] o reconocimiento en vídeo [43]. Este detector tradicional supone un buen punto de partida para posteriores evoluciones ya que en sí mismo ya ha probado un buen comportamiento ante rotaciones y cambios de escala[44]. Está basado en la matriz del segundo momento o matriz de autocorrelación que describe el cambio de intensidad en el vecindario local de un punto $p = (x, y)$, basándose en la

suposición de que en estructuras de tipo esquina (como se definió en 2.1.2) la intensidad de la imagen variará en múltiples direcciones. La matriz M sería 4.4, dondeI_x e I_y son las derivadas de la función de intensidad de la imagen en dicho punto.

$$M = g(x, \upsilon, \sigma) * \begin{bmatrix} I_x^2(x) & I_x I_y(x) \\ I_x I_y(x) & I_y^2(x) \end{bmatrix} \tag{4.4}$$

Donde típicamente $g(x, \upsilon, \sigma)$ es una gaussiana4.5.

$$g(x, \upsilon, \sigma) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{x^2-y^2}{2\sigma^2}\right)} \tag{4.5}$$

Multi-scale Harris, Harris-Laplace y Harris-Affine.

En estos detectores, propuestos por Micolajczyk y Schmid [19], se requiere en primer lugar una adaptación a escala de la matriz del segundo momento explicada anteriormente según se indica en 4.6.

$$M = \mu(x, \sigma_I, \sigma_D) = \sigma_D^2 g(\sigma_I) * \begin{bmatrix} L_x^2(x, \sigma_D) & L_x L_y(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y^2(x, \sigma_D) \end{bmatrix} \tag{4.6}$$

Donde $L(x)$ es un suavizado gaussiano de la imagen mediante un kernel gaussiano $g(\sigma_I)$ de escala σ_I (que determina la escala actual) y por su parte $L_x(x, \sigma_D)$ y $L_y(x, \sigma_D)$ son las derivadas en su respectiva dirección aplicadas a la imagen suavizada en una escala denominada escala de diferenciación σ_D.

Los puntos de interés se localizarían en los máximos locales de una medida, denominada *cornerness*, que se define 4.7.

$$cornerness = det(\mu(x, \sigma_I, \sigma_D)) - \alpha \cdot trace^2(\mu(x, \sigma_I, \sigma_D)) \tag{4.7}$$

Una vez hecha esta adaptación al espacio escala, el método Multi-scale Harris propuesto por Mikolajczyk y Schmid realiza una búsqueda de puntos de interés sobre un set predefinido de escalas $\sigma_n = \xi^n \sigma_0$ [19] (en el que el factor de escala elegido fue$\xi = 1{,}4$ en este caso). Para cada nivel de escala a su vez se detectan los máximos locales para cada punto x en un vecindario de 8 píxeles, para posteriormente desechar mediante un umbral los máximos locales que obtienen menor *cornerness*, ya que serían los menos estables ante cambios en las condiciones de la imagen.

El método Harris-Laplace se sirve del detector de Harris adaptado a escala para localizar puntos en el espacio-escala para, a continuación, como se ha comentado, llevar a cabo una búsqueda iterativa de la escala característica de cada punto mediante el operador Laplaciano-Gaussiano (LoG). De esta manera la escala característica se

encontrará en los máximos locales de la función normalizada 4.8.

$$|LoG(x, \sigma_n)| = \sigma_n^2 \, |L_x x(x, \sigma_n) + L_y y(y, \sigma_n)| \tag{4.8}$$

Mediante un algoritmo iterativo se obtienen puntos de interés donde se maximiza medida de *cornerness* de Harris sobre su vecindario de píxeles (selección espacial) y también se maximiza la función anterior (LoG) en la escala característica del punto (selección de escala).

Junto con los anteriores detectores, fue propuesta una tercera evolución con el objetivo de detectar puntos invariantes a transformaciones afines y cambios de punto de vista. El detector Harris-Affine realiza una normalización afín de los puntos mediante un algoritmo de adaptación de contorno llegando así al algoritmo expuesto en [19].

4.1.1.2. Métodos de matriz hessiana

Hessian detector.

Este detector propuesto por Beaudet [10], hace uso de la matriz hessiana, es decir las derivadas parciales de segundo orden, de la función de intensidad de la imagen $I(x)$. Para una localización $x = (x, y)$ la matriz hessiana sería 4.9.

$$H = \begin{bmatrix} I_{xx}(x) & I_{xy}(x) \\ I_{xy}(x) & I_{yy}(x) \end{bmatrix} \tag{4.9}$$

Este detector se basa en que calculando sobre una imagen la respuesta del determinante, $I_{xx}(x)I_{yy} - I_{xy}^2$, se obtienen máximos locales en regiones y estructuras tipo *blob* (como se definió en 2.1.2).

En este proyecto evaluaremos dos modificaciones de este detector que se tratan, equivalentemente a las del detector de Harris, de una adaptación multi-escala del mismo y otra para la selección automática de la escala característica de las detecciones basada en la Laplaciana de la Gaussiana. Estos detectores son Multi-scale Hessian y Hessian-Laplace.

Multi-scale Hessian, Hessian-Laplace y Hessian Affine.

Estos tres detectores resultan tener una arquitectura idéntica a los detectores homólogos de Harris con la salvedad de que en este caso se parte del determinante de la matriz Hessiana para la extracción de los puntos de interés, que en este caso serían regiones tipo *blob*.

Una de las ventajas de los detectores basados en la matriz hessiana es que, típicamente, extraen un gran número de puntos de interés, dando como resultado una

buena cobertura de los mismos por toda la imagen. Igualmente que en los detectores de Harris, el número de regiones encontradas puede ser controlado, generalmente sesgando el número de puntos, fijando un umbral (*threshold*) para la respuesta del determinante Hessiano o del laplaciano en su caso.

SURF.

El detector de SURF (*Speed-Up Robust Features*) [1] presenta altas similitudes con el detector de SIFT pero mejora significativamente los tiempos de procesado de su predecesor con resultados similares. Esto hace que sus aplicaciones se centren más en técnicas relacionadas con el vídeo gracias a su ventaja en el tiempo de procesado.

Se trata de un detector invariante a escala que se basa en la matriz Hessiana para obtener la localización y la escala de los puntos. La matriz Hessiana se aproxima mediante el uso de un conjunto de filtros de tipo caja sobre imágenes integrales [45] por lo que no es necesario aplicar suavizados gaussianos de una escala a otra.

Una vez que la imagen integral ha sido computada, el cálculo de las intensidades sobre cualquier área rectangular, independientemente del tamaño, se realiza mediante tan sólo 4 sumas. Posteriormente, mediante el uso de los filtros tipo caja, se llega a una aproximación de la respuesta de lo que sería el determinante de la matriz Hessiana para cada punto por un muy bajo coste computacional, ver Figura 4.1.

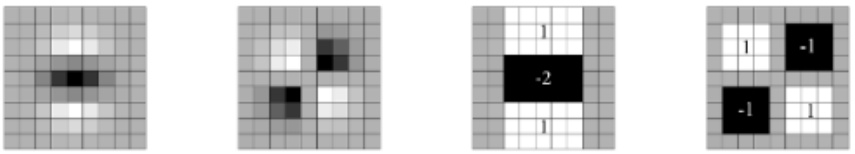


Figura 4.1: De izquierda a derecha, derivadas parciales de Gaussianas en dirección y y dirección xy, y las aproximaciones de SURF usando filtros tipo caja. Fuente [1]

Por último se seleccionan los máximos locales en un vecindario de $3 \times 3 \times 3$ en el espacio-escala mediante una interpolación cuadrática.

Gracias a su gran ventaja en el tiempo de procesado y su buen rendimiento, se utiliza comúnmente en aplicaciones y técnicas relacionadas con el vídeo.

4.1.1.3. Métodos de matriz laplaciana

Detector de SIFT.

El popular algoritmo SIFT (Scale-invariant feature transform) desarrollado por D.

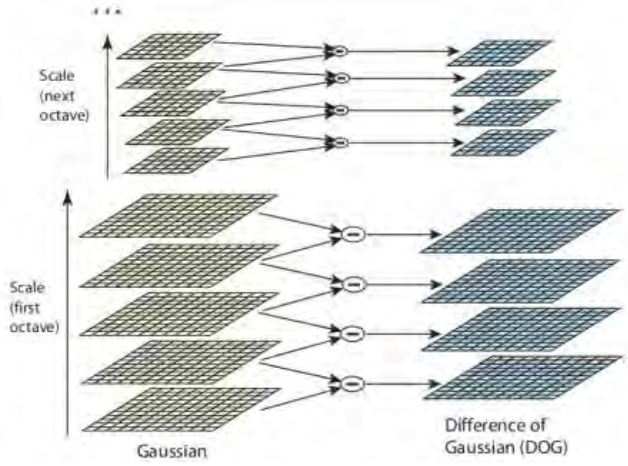


Figura 4.2: Esquema gráfico de la aplicación de la técnica de DOG sobre una imagen. A la derecha las imágenes con el filtrado gaussiano, a la izquierda las imágenes diferencia. Fuente [2].

Lowe [2] combina tanto detección como descripción de puntos de interés. El algoritmo extrae los puntos de interés mediante un método basado en la construcción de pirámides gaussianas (reducción progresiva del tamaño de la imagen) en el espacio-escala, calculando diferencias gaussianas (DoG) entre la escala de cada punto y sus escalas próximas.

En esta primera etapa del algoritmo, se hace uso de la función DoG que busca obtener máximos y mínimos relativos de las diferencias de escalas consecutivas 4.10.

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{4.10}$$

Donde k es la constante multiplicativa de dichas escalas. Esto supone una ventaja desde un punto de vista computacional ya que el cálculo de D se realiza mediante una resta. Otra de las ventajas de la utilización de esta función es que se puede aproximar al Laplaciano de la Gaussiana de escala normalizada, cuyos máximos y mínimos producen puntos de interés más estables que otras funciones como el Gradiente, el Hessiano o el detector de esquinas de Harris.

Para limitar la redundancia en la pirámide de escalas se lleva a cabo un muestreo sucesivo que comienza sobre una imagen expandida con respecto a la original. Cada una de las escalas (u octavas) se divide a su vez en un número entero de escalas con un valor de σ del doble con respecto a la imagen que le precede, lo que para un número reducido de escalas supone bajo coste computacional. El esquema sería el de la Figura 4.2 según el autor[2].

Sobre esta construcción, los máximos y mínimos locales se obtienen mediante la comparación de un píxel con sus vecinos en su misma escala y con los de las escalas contiguas como se muestra en la Figura 4.3, donde el píxel examinado está marcado con una X.

Una vez obtenidos los máximos y mínimos candidatos a ser puntos clave se someten a un proceso en el que, según unos criterios, los menos estables son descartados.

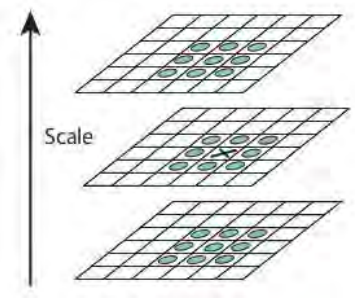


Figura 4.3: Esquema gráfico de la aplicación de mínimos y máximos locales en vecindario con escalas contiguas. Fuente [2].

Por un lado se aplica un proceso de umbralización para descartar puntos con bajo contraste que serían los menos robustos ante cambios de iluminación. Otros puntos que también son descartados son los que se sitúan en bordes difusos (para los que la respuesta de DoG produce extremos relativos), por medio de una propiedad de la matriz hessiana que permite discriminar los que son vulnerables ante cambios de punto de vista.

Una vez descartados los puntos inestables, al resto de puntos se les asigna una orientación basada en las propiedades locales de la imagen en cada punto para conseguir invariancia a rotación. El algoritmo además contempla que haya puntos con más de una orientación principal, lo que se traduce en una mayor estabilidad de estos. Tras asignar las orientaciones comienza la etapa de descripción que describiremos en secciones posteriores.

D1 de Lindeberg.

Se trata de un detector basado en la publicación de Lindeberg [46], en la que propone dos formas de detectar los puntos de interés en el espacio escala, el operador Laplaciano 4.11.

$$\nabla^2_{norm}L = t(L_{xx} + L_{yy}) \tag{4.11}$$

O el determinante de la matriz Hessiana 4.12.

$$detH_{norm}L = t^2(L_{xx}L_{yy} - L^2_{xy}) \tag{4.12}$$

Para después extraer los puntos en base a una medida de fuerza según 4.13.

$$\mathcal{D}_{1,norm}L = \begin{cases} t^2(det\mathcal{H}L - k\,trace^2\mathcal{H}L) & det\mathcal{H}L - k\,trace^2\mathcal{H}L > 0 \\ 0 & otherwise \end{cases} \tag{4.13}$$

Donde $\mathcal{H}L$ es la matriz Hessiana calculada en cada punto de la imagen y t es la escala actual.

4.1.2. Detectores de entorno.

El algoritmo precursor de estas técnicas es FAST. Todas las restantes surgen a partir o en base a ella, por lo que se presentarán secuencialmente, y por lo que carece de sentido incluir una nueva categorización. Aparte de FAST, en esta categoría se van a describir AGAST, BRISK y ORB.

4.1.2.1. Detector de FAST.

FAST (*Features from Accelerated Segment Test*) [3], publicado en 2005, es un algoritmo de detección de esquinas basado en el detector SUSAN (*Smallest Univalue Segment Assimilating Nucleus*) [47]. Estos dos algoritmos llevan a cabo una comparación de las intensidades de un píxel central, candidato a punto de interés, con píxeles de su vecindario. SUSAN evalúa una fracción píxeles en el vecindario del píxel central que calcula en función de que tengan una intensidad similar a la de éste. FAST lleva la idea más allá y evalúa únicamente cambios de intensidad en los píxeles en un círculo con un radio fijado alrededor del punto (4.4).

Una versión muy utilizada de FAST es FAST 9-16, que implica que la intensidad de 9 de los 16 píxeles de un círculo de radio fijo, debe ser menor o mayor que la intensidad del píxel central según un umbral, ver ejemplo en Figura 4.4.

Los primeros píxeles evaluados en FAST 9-16 son 1, 5, 9 y 13 según el algoritmo. Si el píxel C es el centro de una esquina, por lo menos tres de esos píxeles deben ser

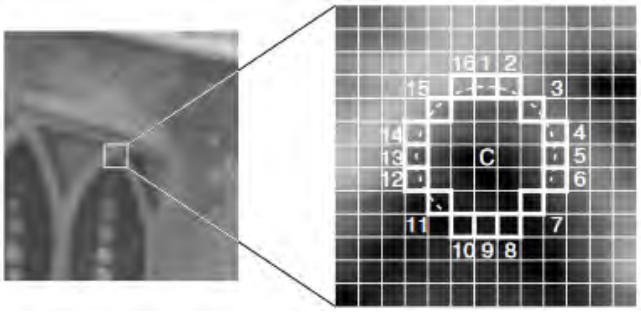


Figura 4.4: Esquema de la técnica de detección de BRISK en el espacio-escala. Fuente [3]

más brillantes o más oscuros que el píxel C.

Este algoritmo de detección de esquinas está presente sobretodo en aplicaciones de vídeo debido a su alta velocidad sin, según presenta el autor, penalizar en repetibilidad.

4.1.2.2. Detector de AGAST.

Como FAST, se trata de un algoritmo de detección basado en el Accelerated Segment Test propuesto por E. Mair y otros [4] con el objetivo de mejorar en velocidad y prestaciones a su predecesor FAST.

Según sus autores, la motivación de AGAST (*Adaptative and Generic Accelerated Segment Test*) es que el árbol de decisión de FAST tiene que ser aprendido en cada nuevo entorno desde cero y se basa en un árbol de decisión ternario, mientras que uno binario sería más eficiente computacionalmente. Además encuentran que el detector de FAST necesita reaprender la configuración del entorno de un píxel esquina si la cámara cambia de vista y especialmente si hay rotación.

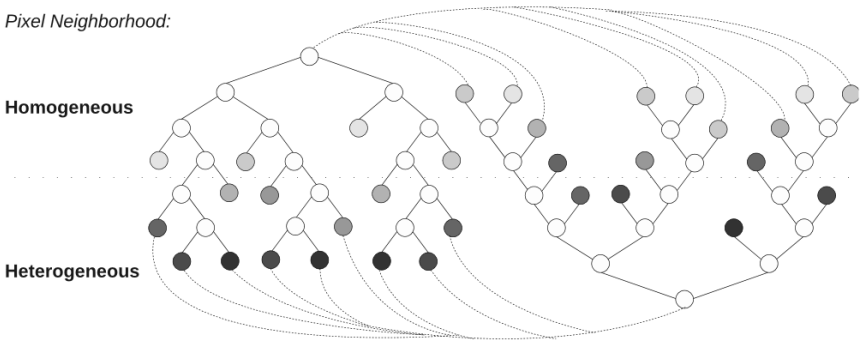


Figura 4.5: Esquema del árbol de detección de AGAST. Fuente [4].

Para lidiar con estos aspectos, proponen un detector de esquinas que se computa mediante un árbol binario de decisión genérico, ver Figura 4.5. Además combinando dos árboles en uno, el detector de esquinas decide en primera instancia si el vecindario del píxel se trata de un entorno homogéneo o heterogéneo, lo que, según los autores, es más rápido que el entrenamiento de FAST.

El árbol de la izquierda logra menos evaluaciones de píxeles (rutas de decisión más cortas) en un vecindario de píxeles homogéneo, mientras que el de la derecha está optimizado para las regiones texturadas.

4.1.2.3. Detector de BRISK.

El método de BRISK, presentado en 2011 [5], consta tanto de detector como de descriptor de puntos de interés. El método de detección está basado en el detector AGAST [4], que a su vez está basado en el detector FAST [3]. Con el objetivo de conseguir invariancia a escala, van un paso más allá del detector FAST y buscan máximos relativos utilizando la medida de *saliencia*(s) de FAST, no solo en una imagen plana sino también en un espacio-escala, para lo que además se calcula la escala característica de cada *keypoint*.

Las capas de la pirámide del espacio-escala consisten típicamente en n octavas c_i y n intraoctavas d_i, para $i = 0, 1, \ldots, n-1$ y típicamente $n = 4$. Las octavas están formadas progresivamente por un muestreo mitad de la imagen anterior, donde c_0 es la imagen original, como se muestra en el esquema de la Figura 4.6.

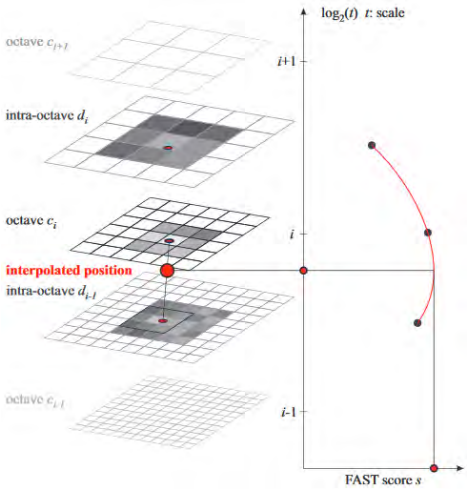


Figura 4.6: Esquema de la técnica de detección de BRISK en el espacio-escala. Fuente [5]

Cada punto analizado se compara con un umbral para determinar si es máximo en su vecindario (3x3) en su misma capa y en las dos adyacentes.

4.1.2.4. Detector de ORB.

ORB (*Oriented FAST and Rotated Brief*) es un método que consta de detección y descripción de puntos de interés propuesto por E. Rublee [17]. Se trata básicamente de la fusión del detector de FAST con el descriptor de BRIEF con una variedad de modificaciones para conseguir mejores prestaciones y eficiencia. En el caso del detector, la principal aportación del método es una asignación de una componente de orientación para conseguir mayor invariancia rotacional.

El detector, que denotan por oFAST comienza detectando puntos FAST que toman como parámetros la diferencia de intensidad del píxel central con aquellos que hay en un anillo circular alrededor del centro (en este caso 9 píxeles). A continuación sobre estos puntos se asigna una medida de *cornerness* de Harris, ya que apuntan que FAST tiene una gran respuesta también para regiones tipo borde, para posteriormente seleccionar los que obtienen la mejor respuesta. Además emplean una pirámide de escalas para extraer puntos FAST en cada nivel de la pirámide.

Para obtener la orientación de cada esquina detectada, utilizan como medida el centroide de intensidad, que está desplazado del centro del segmento AST, con lo que posteriormente es posible calcular su orientación. Rosten [3], definía el momento de cada *patch* como 4.14.

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{4.14}$$

Y con este momento el centroide se calcularía como 4.15.

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{4.15}$$

Con lo que construyen un vector desde el centro de la esquina O, hasta el centroide OC. La orientación del patch sería entonces calculada mediante el arcotangente de los dos primeros momentos (teniendo en cuenta el cuadrante) 4.16.

$$\theta = atan2(m_{01}, m_{10}) \tag{4.16}$$

4.2. Evaluación comparativa teórica.

Una vez analizadas todas las técnicas de detección consideradas de interés para la evaluación comparativa, es posible llevar a cabo un análisis teórico de las propiedades que cada uno de los algoritmos presenta.

Para ello, se han fijado una serie de características a valorar en dicha comparación. Dichas características han sido:

- Categoría inicial en la que se han catalogado, esto es, en función de la técnica de detección, diferenciando si se estima mediante un operador matemático (*Math*), o con una técnica de detección por entorno (*Ent*).
- Si la técnica implementa la generación de un espacio escala (E-S) para aportar invariancias a las detecciones.
- Tipo de punto de interés, haciendo referencia a la terminología definida en la subsección 2.1.2, diferenciando entre esquina, borde y *blob*.
- Invariancias teóricas, siendo las tres principales, y por tanto las escogidas, a escala, a punto de vista (*viewpoint*) y a rotación.
- Otras más individuales como técnicas precursoras y mejoras sobre ella, fortalezas u operadores clave.

Tras un exhaustivo análisis teórico de todas las características mencionadas, se ha recogido la información en la Tabla 4.1. Para una mejor comprensión de la misma, aportar dos notas aclaratorias:

Valoraciones de las columnas Categoría, E-S y Puntos de interés:

- Un punto (●) indica afirmativo en esa propiedad o categoría, y su ausencia indica lo contrario.

Valoraciones de la columna Implementa Invariancia:

- El valor (+++) indica que el algoritmo obtiene unos resultados claramente mejores que el resto de algoritmos (pudiendo ser más de un algoritmo los que obtienen este resultado).
- El valor (++) indica un valor intermedio en comparación con otros algoritmos
- El valor (+) indica unos resultados claramente peores que el resto de algoritmos (pudiendo ser más de un algoritmo los que obtienen este resultado).

Detector	*Año*	*Categoría*		*E-S*	*Puntos de interés*			*Implementa Invariancia*			*Fortalezas*	*Precursor*	*Mejora precursor*	*Operador clave*
		Math	*Ent*		*Esquina*	*Borde*	*Blob*	*Escala*	*Viewpoint*	*Rotación*				
Harris	*1988*	●			●	●						-	-	*Matriz de covarianza*
Hessian	*1978*	●					●				*Cantidad*	-	-	*Matriz Hessiana*
Multi-scale Harris	*2004*	●		●	●	●		+				*Harris*	*E-S*	*Matriz de covarianza*
Multi-scale Hessian	*2004*	●		●			●	+			*Cantidad*	*Hessian*	*E-S*	*Matriz Hessiana*
Harris Laplace	*2004*	●		●	●	●		++		++		*Harris*	*E-S*	*Precursor + LoG*
Hessian Laplace	*2004*	●		●			●	++		++	*Cantidad*	*Hessian*	*E-S*	*Precursor + LoG*
SIFT	*1999*	●		●			●	+++	*++*	*++*	*Robustez*	-	-	*DoG*
SURF	*2006*	●		●			●	*++*	++	++	*Eficiencia*	*SIFT*	*Box filters*	*Matriz Hessiana*
D1	*2013*	●		●			●	+++	+		*Robustez*	*DoG*		*Matriz Laplaciana*
BRISK	*2011*		●	●	●			++	+		*Eficiencia*	*AGAST*	*E-S*	
FAST	*2005*		●		●				+	+	*Eficiencia*	*SUSAN*	*AST*	*AST*
AGAST	*2010*		●		●			+	++	+++	*Eficiencia*	*FAST*	*ABBA**	*ABBA**
ORB	*2011*		●	●	●			+	++	+++	*Eficiencia*	*FAST*	*E-S y orientación*	*Centroides*

Tabla 4.1: Análisis comparativo teórico de las técnicas de detección. Un punto (●) indica afirmativo. Un (+) indica baja invariancia y (+++) indica alta invariancia. De izquierda a derecha se ven los detectores y el año en el que fueron propuestos. El tipo de detección que realizan, el tipo de puntos de interés que son y las invariancias que implementan. Las últimas columnas atienden a criterios de fortalezas principal del método, precursores y qué aportaron sobre ellos, y su operador clave.*ABBA (árbol de búsqueda binario adaptativo).

4.3. Evaluación y análisis.

Como uno de los objetivos principales del proyecto, a continuación de la evaluación teórica presentada, se llevará a cabo una evaluación práctica sobre el nuevo marco común y mejorado aportado en este proyecto.

La evaluación se realizará en primer lugar sobre el conjunto de datos presentado por K. Mikolajczyk [18] como principal referencia en el estado del arte y en segundo lugar sobre el nuevo conjunto de datos construidos para el nuevo marco de evaluación de este proyecto.

Los datos se calcularán y presentarán en base a las métricas expuestas en la sección3.2, que a su vez permitirán analizar resultados y comparar con las conclusiones extraídas de la evaluación teórica.

4.3.1. Evaluación sobre el conjunto de datos de K. Mikolajczyk

El dataset consta de ocho secuencias de seis imágenes cada una, de manera que la primera imagen de cada secuencia es sobre la que se comparan las cinco restantes.

En siete de las secuencias hay movimientos de cámara por lo que para todas ellas ha sido necesario obtener los *groundtruths* en base a las homografías aportadas por el autor. Dichos movimientos, en unos casos son producto de las transformaciones geométricas que presenta la secuencia y en otros casos se trata de pequeños movimientos de la cámara a la hora de capturar las imágenes, pero que igualmente hacen necesario el cálculo del *groundtruth* a partir de las homografías.

A continuación se expondrán y analizarán los resultados obtenidos por cada detector para cada una de las secuencias de las que consta el dataset. Se incluirá en primer lugar una descripción las características relevantes de la secuencia para esclarecer y contextualizar el análisis.

Las transformaciones de las ocho secuencias son:

- Dos secuencias con blur
- Dos secuencias de rotación combinado con zoom
- Dos secuencias con cambio de punto de vista
- Una secuencia de cambio de iluminación
- Una secuencia con compresión JPEG

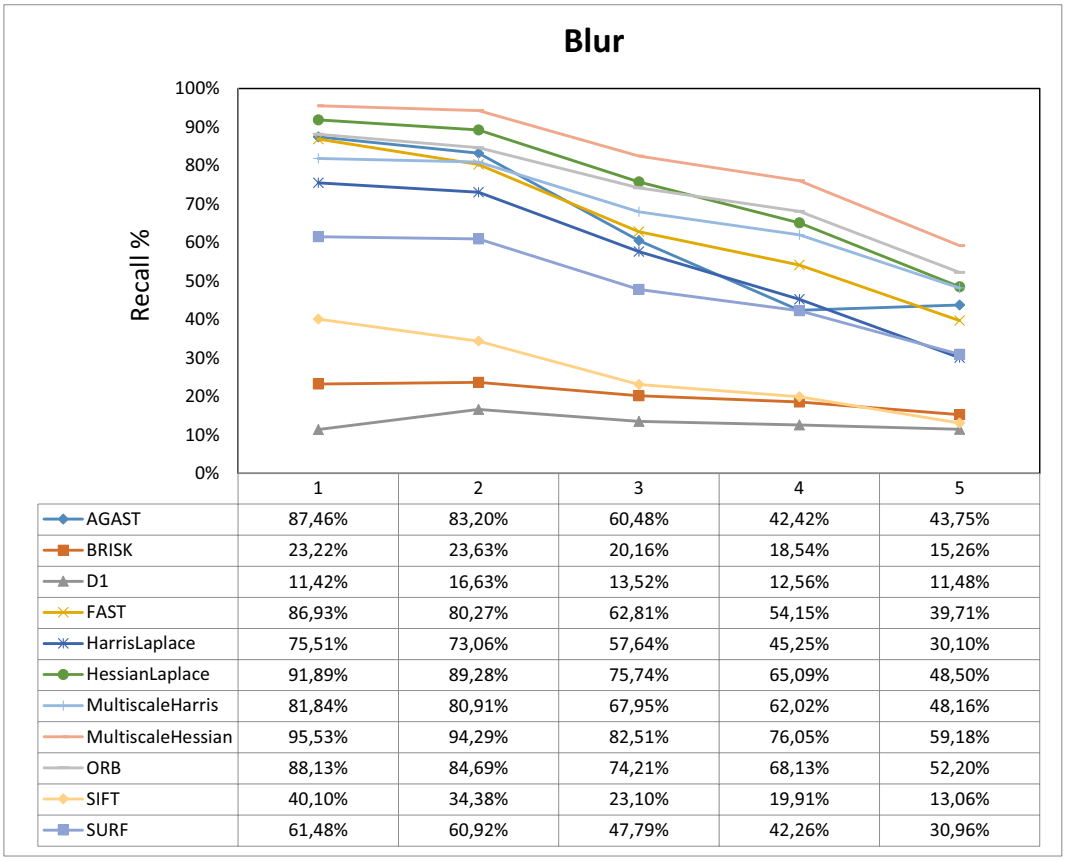
En el caso de las transformaciones de blur, rotación y zoom y cambio de punto de vista, cada una de las dos secuencias presenta unas condiciones de escena diferentes:

en una de ellas la escena contiene regiones homogéneas con bordes distinguibles y en la otra, la escena presenta texturas repetitivas con diferentes formas.

4.3.1.1. Blur sobre escena con regiones homogéneas

La secuencia parte de una imagen de referencia nítida sobre la que se va produciendo una borrosidad progresiva en las siguientes imágenes.

Los resultados comparativos de recall para las cinco imágenes y los 11 algoritmos son los siguientes:



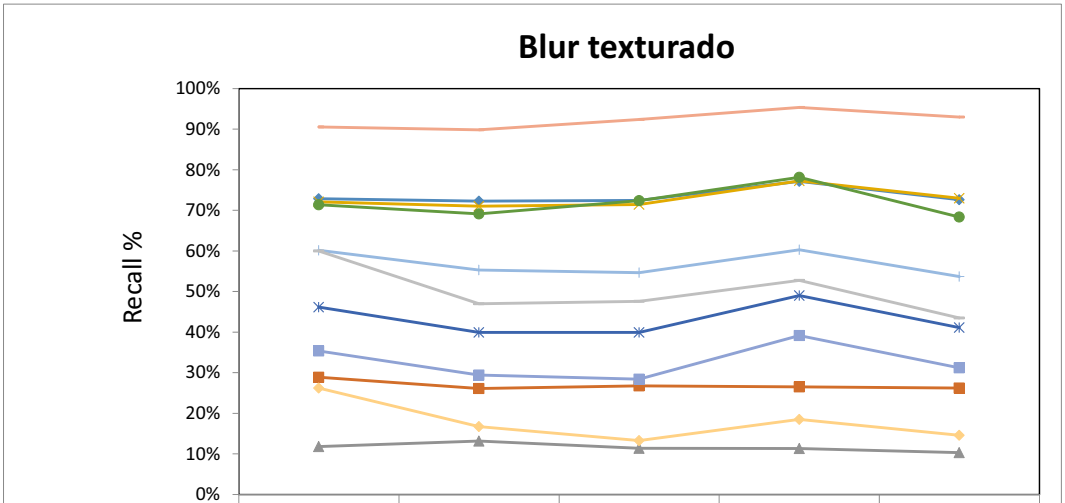
	1	2	3	4	5
AGAST	87,46%	83,20%	60,48%	42,42%	43,75%
BRISK	23,22%	23,63%	20,16%	18,54%	15,26%
D1	11,42%	16,63%	13,52%	12,56%	11,48%
FAST	86,93%	80,27%	62,81%	54,15%	39,71%
HarrisLaplace	75,51%	73,06%	57,64%	45,25%	30,10%
HessianLaplace	91,89%	89,28%	75,74%	65,09%	48,50%
MultiscaleHarris	81,84%	80,91%	67,95%	62,02%	48,16%
MultiscaleHessian	95,53%	94,29%	82,51%	76,05%	59,18%
ORB	88,13%	84,69%	74,21%	68,13%	52,20%
SIFT	40,10%	34,38%	23,10%	19,91%	13,06%
SURF	61,48%	60,92%	47,79%	42,26%	30,96%

Figura 4.7: Comparativa de detectores del estado del arte frente a la propiedad de blur

En ésta primera secuencia hay tres algoritmos que obtienen unos valores de recall notablemente por debajo de los ocho restantes, que son BRISK, SIFT y D1, sin embargo se puede ver también como la pendiente no desciende con tanta claridad como en el caso de otros algoritmos. Ésta tendencia también se observará en algunas secuencias posteriores.

4.3.1.2. Blur sobre escena texturada

La imagen original presenta una escena muy texturada que se va emborronando progresivamente.



	1	2	3	4	5
AGAST	72,83%	72,27%	72,47%	77,12%	72,60%
BRISK	28,88%	26,11%	26,74%	26,50%	26,23%
D1	11,82%	13,14%	11,37%	11,29%	10,33%
FAST	72,10%	71,01%	71,45%	77,23%	72,99%
HarrisLaplace	46,12%	39,92%	39,94%	49,00%	41,09%
HessianLaplace	71,36%	69,11%	72,35%	78,11%	68,33%
MultiscaleHarris	60,14%	55,29%	54,64%	60,29%	53,69%
MultiscaleHessian	90,57%	89,80%	92,40%	95,34%	92,99%
ORB	59,99%	47,00%	47,56%	52,76%	43,50%
SIFT	26,22%	16,71%	13,30%	18,52%	14,60%
SURF	35,36%	29,41%	28,38%	39,16%	31,20%

Figura 4.8: Recall secuencia de blur con texturado

Como se puede observar, el algoritmo Multi-scale Hessian obtiene muy buenos resultados en toda la secuencia, prácticamente sin importar el nivel de degradación. Esto es debido a que se trata de un detector de blobs que es la textura predominante en la imagen, lo que provoca que se detecten gran cantidad de puntos en todas las imágenes de la secuencia.

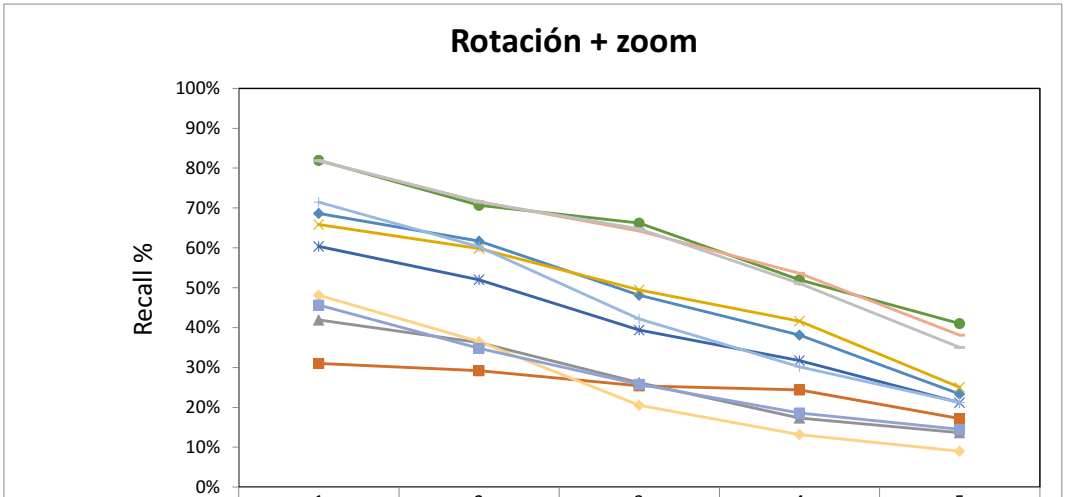
Como se puede ver no se produce un descenso del recall de los algoritmos conforme aumenta la borrosidad (como ocurría en la secuencia anterior), esto es a causa de las características texturadas de la imagen, que se traducen en una detección de puntos de interés para todas las secuencias, que no se ve reducido por el *blurring*.

Se puede concluir que este tipo de escenas texturadas no son indicadas para medir la respuesta de los algoritmos a una transformación de este tipo, ya que lo que se

consigue en su lugar es medir la respuesta de éstos ante la textura de la escena, quedando enmascarada la respuesta ante la degradación por blur.

4.3.1.3. Rotación combinado con zoom

La secuencia parte de de una escena que en las sucesivas imágenes se va alejando y rotando en ángulos que no superan los 90 grados (ver 3.1.2.6).



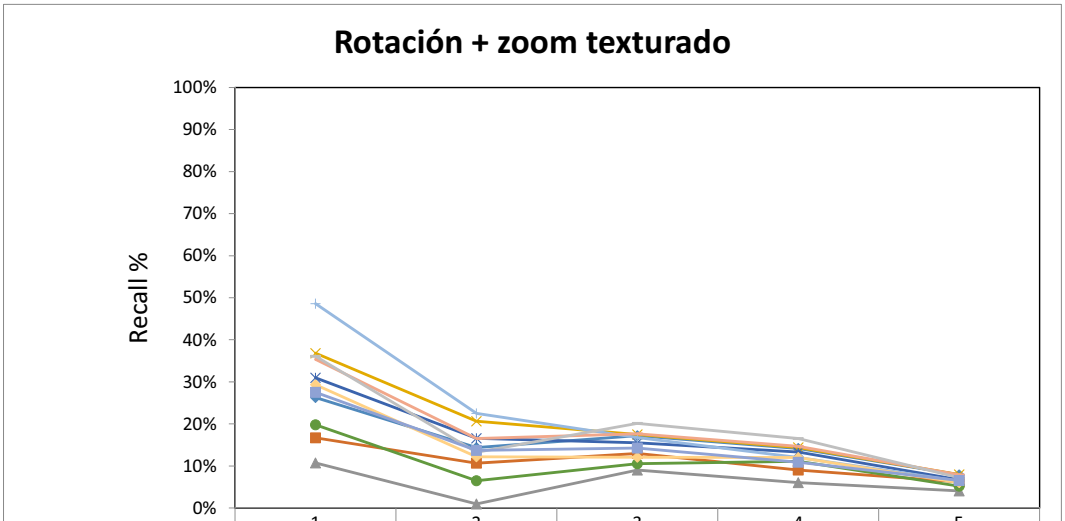
	1	2	3	4	5
AGAST	68,62%	61,69%	48,16%	38,11%	23,37%
BRISK	31,02%	29,21%	25,43%	24,36%	17,23%
D1	41,89%	36,23%	26,18%	17,31%	13,62%
FAST	65,87%	59,83%	49,44%	41,59%	25,03%
HarrisLaplace	60,36%	51,97%	39,39%	31,76%	21,18%
HessianLaplace	81,97%	70,71%	66,22%	51,98%	41,00%
MultiscaleHarris	71,47%	60,26%	42,15%	30,15%	21,26%
MultiscaleHessian	81,80%	71,57%	64,22%	53,63%	38,10%
ORB	81,89%	71,53%	64,76%	51,07%	35,06%
SIFT	48,13%	36,48%	20,55%	13,16%	8,98%
SURF	45,65%	34,80%	25,77%	18,56%	14,46%

Figura 4.9: Recall secuencia de escala combinado con rotación

Los resultados muestran tres algoritmos que sobresalen sobre los demás, ORB, AGAST y Multi-scale Hessian. Otro efecto observable es la pendiente descendente que muestran los algoritmos salvo en el caso de BRISK, de lo que se puede deducir que parte de los puntos detectados serían muy robustos a este tipo de transformaciones. Esto confirmaría el éxito de las estrategias que implementa este algoritmo para hacer frente a este tipo de transformaciones, apoyándose también en el buen funcionamiento de su predecesor, AGAST, en esta secuencia.

4.3.1.4. Rotación combinado con zoom sobre escena texturada

La secuencia parte de una escena texturada que se va alejando y rotando en ángulos de hasta 180 grados progresivamente.



	1	2	3	4	5
AGAST	26,27%	14,38%	17,22%	14,13%	7,89%
BRISK	16,72%	10,68%	12,96%	9,03%	6,19%
D1	10,70%	0,96%	9,02%	6,06%	4,07%
FAST	36,82%	20,66%	17,49%	14,39%	7,96%
HarrisLaplace	30,97%	16,57%	15,53%	13,33%	6,74%
HessianLaplace	19,81%	6,49%	10,54%	11,14%	5,20%
MultiscaleHarris	48,58%	22,51%	16,77%	11,99%	6,45%
MultiscaleHessian	35,36%	16,55%	17,63%	14,66%	7,91%
ORB	36,07%	13,24%	20,11%	16,54%	7,21%
SIFT	29,32%	12,16%	12,09%	12,01%	6,03%
SURF	27,54%	13,69%	14,25%	10,87%	6,54%

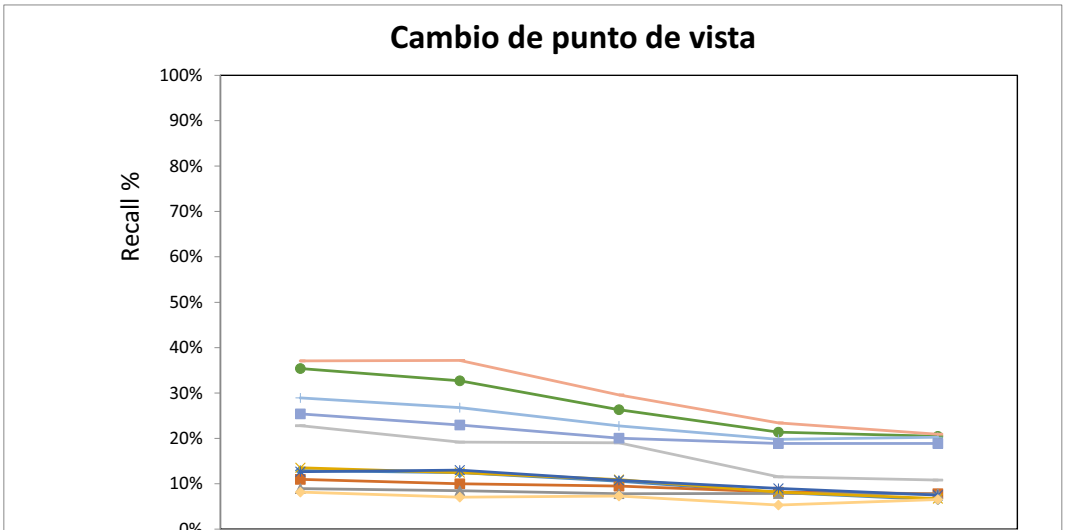
Figura 4.10: Recall secuencia zoom y rotación sobre texturado

Los resultados que obtienen los algoritmos sobre esta secuencia son, en todos los casos, muy pobres. Si se observan con más detalle las imágenes se pueden extraer varias causas de éstos malos resultados: En primer lugar, las imágenes son de muy mala calidad, lo que hace que la estructura que se observaba en la primera imagen es casi inapreciable a simple vista a partir de el segundo paso de zoom. Además aparecen otras transformaciones en la imagen que no son objeto de estudio, como falta de nitidez y cambios de iluminación.

Por estos motivos, no se trata de una secuencia adecuada para estudiar el rendimiento de los algoritmos ante efectos de escalado y rotación y la secuencia será excluida de la evaluación de descriptores.

4.3.1.5. Cambio de punto de vista

La escena parte de una posición frontal de un dibujo plano y termina con una imagen del dibujo desde una posición lateral de unos 60 grados pasando por pasos en los que se introducen a la vez rotaciones importantes de la cámara.



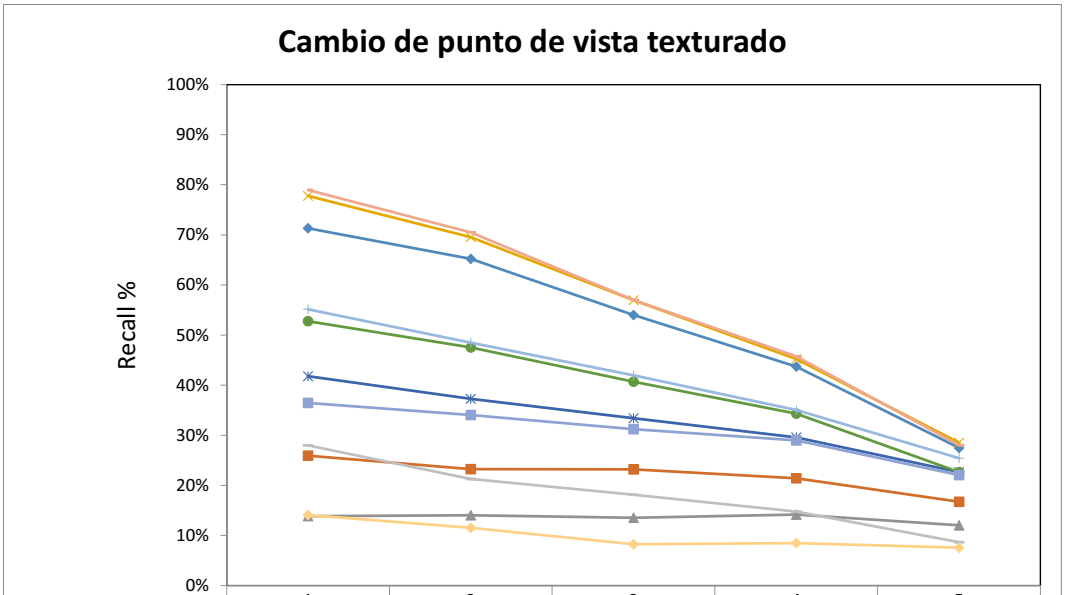
	1	2	3	4	5
AGAST	12,90%	12,47%	10,57%	8,08%	6,50%
BRISK	10,96%	9,99%	9,49%	8,18%	7,82%
D1	8,95%	8,46%	7,80%	7,88%	7,75%
FAST	13,54%	12,47%	10,93%	8,18%	6,75%
HarrisLaplace	12,68%	13,01%	10,78%	8,99%	7,50%
HessianLaplace	35,42%	32,69%	26,33%	21,39%	20,43%
MultiscaleHarris	28,90%	26,76%	22,73%	19,77%	20,27%
MultiscaleHessian	37,10%	37,17%	29,59%	23,40%	20,93%
ORB	22,81%	19,20%	19,03%	11,54%	10,81%
SIFT	8,19%	7,07%	7,33%	5,30%	6,55%
SURF	25,39%	22,94%	20,05%	18,90%	18,88%

Figura 4.11: Recall secuencia de cambio de punto de vista

Según los datos del recall de los algoritmos, casi todos consiguen malos resultados para esta secuencia en comparación con otras transformaciones, pero especialmente algunos algoritmos basados en FAST tienen una respuesta que aunque no decae paso tras paso, sí que se puede considerar que consiguen poco porcentaje de puntos robustos ante este tipo de transformaciones.

4.3.1.6. Cambio de punto de vista sobre escena texturada

La secuencia, de igual manera que la anterior, parte de una posición frontal de una escena texturada y llega hasta una posición lateral de unos 60 grados respecto a



	1	2	3	4	5
AGAST	71,34%	65,20%	54,02%	43,71%	27,46%
BRISK	25,94%	23,25%	23,21%	21,40%	16,71%
D1	13,85%	14,05%	13,55%	14,19%	12,02%
FAST	77,78%	69,54%	56,98%	45,21%	28,53%
HarrisLaplace	41,77%	37,25%	33,41%	29,55%	22,64%
HessianLaplace	52,75%	47,55%	40,68%	34,30%	22,66%
MultiscaleHarris	55,16%	48,47%	41,96%	35,03%	25,43%
MultiscaleHessian	78,97%	70,48%	57,02%	45,75%	28,02%
ORB	27,96%	21,29%	18,11%	14,76%	8,67%
SIFT	14,06%	11,55%	8,26%	8,48%	7,58%
SURF	36,45%	34,02%	31,22%	28,98%	22,06%

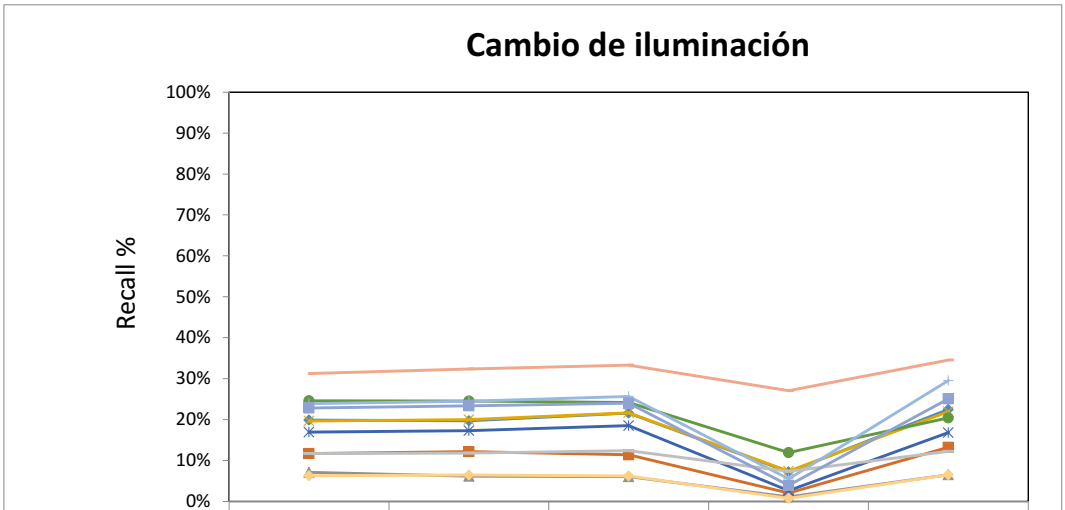
Figura 4.12: Recall secuencia de cambio de punto de vista sobre escena texturada

la original.

A la vista de los resultados se ve que de nuevo hay algoritmos que obtienen buenos resultados para la primera transformación y luego decaen con más velocidad en las siguientes, mientras que hay otros algoritmos que obtienen peores resultados en las primeras secuencias pero se ven menos penalizados en las posteriores. En este segundo caso se puede decir que dichos algoritmos consiguen un bajo porcentaje de buenas detecciones pero éstas son más robustas, por ejemplo BRISK, SURF o D1.

4.3.1.7. Cambio de Iluminación

La escena parte de una imagen de referencia aclarada y continúa con oscurecidos progresivos de la misma escena.



	1	2	3	4	5
AGAST	19,86%	19,67%	21,57%	7,19%	22,48%
BRISK	11,69%	12,17%	11,38%	2,04%	13,22%
D1	7,10%	6,18%	6,02%	1,10%	6,50%
FAST	19,65%	20,01%	21,65%	7,40%	21,93%
HarrisLaplace	16,93%	17,30%	18,50%	2,65%	16,80%
HessianLaplace	24,58%	24,53%	24,19%	11,92%	20,41%
MultiscaleHarris	23,88%	24,44%	25,67%	5,51%	29,50%
MultiscaleHessian	31,24%	32,34%	33,28%	27,06%	34,53%
ORB	11,67%	11,79%	12,41%	7,34%	12,23%
SIFT	6,26%	6,46%	6,22%	0,77%	6,52%
SURF	22,84%	23,37%	23,99%	3,93%	25,07%

Figura 4.13: Recall secuencia de cambio de iluminación

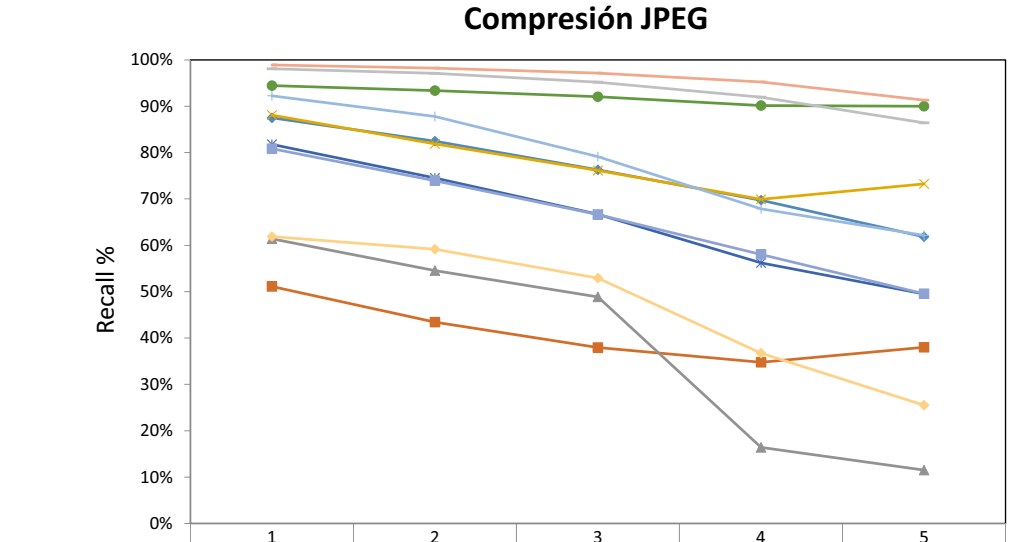
Los malos resultados generalizados de la cuarta transformación (teniendo en cuenta que el oscurecido es menor que en la última foto) sólo se explican por un mal cálculo de la homografía por parte de los autores, ya que según se explica en su web, las homografías fueron calculadas mediante una herramienta de anotación manual, que podría implicar estos errores.

Este suceso pone de manifiesto que realizar las fotografías con los pequeños movimientos de cámara mencionados puede ocasionar errores en la evaluación, ya que el cálculo de las homografías es un proceso cuyo resultado está en función de una anotación manual que puede ser desde imprecisa hasta totalmente errónea (como el caso que se expone).

Este hecho refuerza aún más la construcción del dataset aportado para este proyecto, en el que las fotografías se han tomado en posiciones de cámara fija cuando la transformación no necesita movimientos de cámara.

4.3.1.8. Compresión JPEG

Según explican los autores, esta secuencia se ha generado mediante una herramienta por la que se han variado los parámetros de calidad de la foto desde un 40 % hasta un 2 %. Es la única secuencia del dataset que no requiere del uso de homografías



	1	2	3	4	5
AGAST	87,50%	82,47%	76,28%	69,72%	61,83%
BRISK	51,14%	43,43%	37,95%	34,76%	38,00%
D1	61,42%	54,56%	48,84%	16,42%	11,52%
FAST	88,10%	81,85%	76,15%	69,92%	73,24%
HarrisLaplace	81,74%	74,52%	66,71%	56,18%	49,52%
HessianLaplace	94,43%	93,37%	92,06%	90,15%	90,01%
MultiscaleHarris	92,26%	87,79%	79,14%	67,86%	62,15%
MultiscaleHessian	98,90%	98,22%	97,14%	95,22%	91,31%
ORB	98,06%	97,11%	95,16%	91,97%	86,43%
SIFT	61,89%	59,15%	52,94%	36,71%	25,49%
SURF	80,83%	73,91%	66,63%	58,04%	49,53%

Figura 4.14: Recall secuencia con compresión JPEG

Los resultados para ésta secuencia están en la misma línea que para algunas de las secuencias anteriores en las que algunos algoritmos obtienen buenos resultados para los primeros niveles pero luego decaen con rapidez. Destacan los buenos resultados de ORB, Hessian-Laplace y Multi-scale Hessian.

4.3.1.9. Conclusiones generales

Se han observado algunos fallos importantes en ciertas secuencias del dataset que se tratarán de corregir en la evaluación del dataset elaborado en el marco de evaluación de éste proyecto.

En primer lugar se ha observado que para las secuencias que presentan texturado, los resultados obtenidos obedecen más a la respuesta de los algoritmos ante las propias texturas que a la respuesta a las transformaciones que se desean evaluar. Por lo que no se incluirán este tipo de propiedades en las secuencias que se construyan.

En segundo lugar, se ha observado que el cálculo de homografías puede conllevar errores puesto que implica una parte del proceso manual. Esto a su vez conduce a errores de evaluación. Para la construcción del conjunto de datos que se aportará, se evitarán tener que realizar este proceso cuando no sea necesario, y en el caso de ser necesario se ha adaptado una herramienta para hacerlo de la manera más precisa posible.

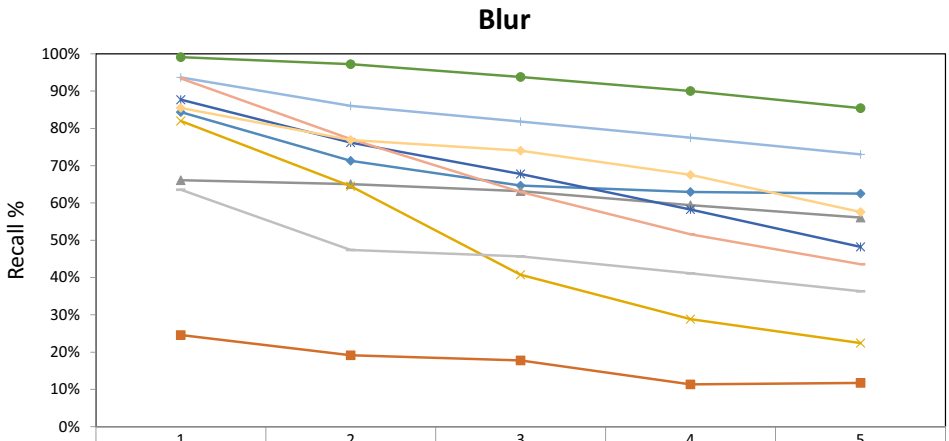
Por último se ha observado que algunos algoritmos obtienen unos resultados demasiado por encima de lo que cabría esperar en comparación con algoritmos más novedosos que deberían superarles. Es el caso del algoritmo Multi-scale Hessian especialmente, y en menor medida los algoritmos de Hessian-Laplace, Harris-Laplace y Multi-scale Harris. Esto es debido a que la evaluación se ha realizado ejecutando estos algoritmos con un parámetro de *threshold* recomendado por los autores que deriva en una detección masiva de puntos por toda la imagen. A causa de esto, en zonas de la imagen texturadas o con abundancia de variaciones de otro tipo, se detectan gran cantidad de puntos de interés a lo largo de todas las imágenes de la secuencia, lo que implica una dificultad para discriminar los TP de los FN que deriva en unos buenos resultados de recall que están inflados por la presencia de estos fenómenos. Además el tiempo de ejecución del algoritmo Multi-scale Hessian es inasumible para ser evaluado sobre un dataset con imágenes de tamaño hasta diez veces superior que las evaluadas hasta ahora, motivos por los se ha descartado este algoritmo de la evaluación sobre el dataset propuesto en este proyecto.

4.3.2. Evaluación sobre el nuevo conjunto de datos aportado

El dataset consta de doce secuencias con las distintas propiedades que se explicaron en la sección 3.1. A continuación se presentarán y analizarán los resultados de recall obtenidos para las distintas secuencias, además de comentar algunas características de las imágenes, cuando las haya, que puedan tener influencia en los resultados de la evaluación.

4.3.2.1. Blur

La mayor parte de la escena escogida para la secuencia corresponde a la fachada de un edificio (predominan esquinas y bordes definidos) y otra parte importante de la imagen corresponde a vegetación y arbolado (apariencia texturada y predominancia de blobs).



	1	2	3	4	5
AGAST	84,40%	71,30%	64,69%	62,96%	62,48%
BRISK	24,59%	19,13%	17,77%	11,34%	11,72%
D1	66,08%	65,05%	63,19%	59,41%	56,07%
FAST	81,97%	64,52%	40,74%	28,85%	22,44%
HarrisLaplace	87,69%	76,15%	67,74%	58,25%	48,24%
HessianLaplace	99,11%	97,19%	93,78%	90,04%	85,43%
MultiscaleHarris	93,66%	86,01%	81,81%	77,49%	73,02%
ORB	93,38%	77,12%	62,96%	51,58%	43,59%
SIFT	63,55%	47,38%	45,65%	41,14%	36,31%
SURF	85,50%	76,89%	74,01%	67,54%	57,56%

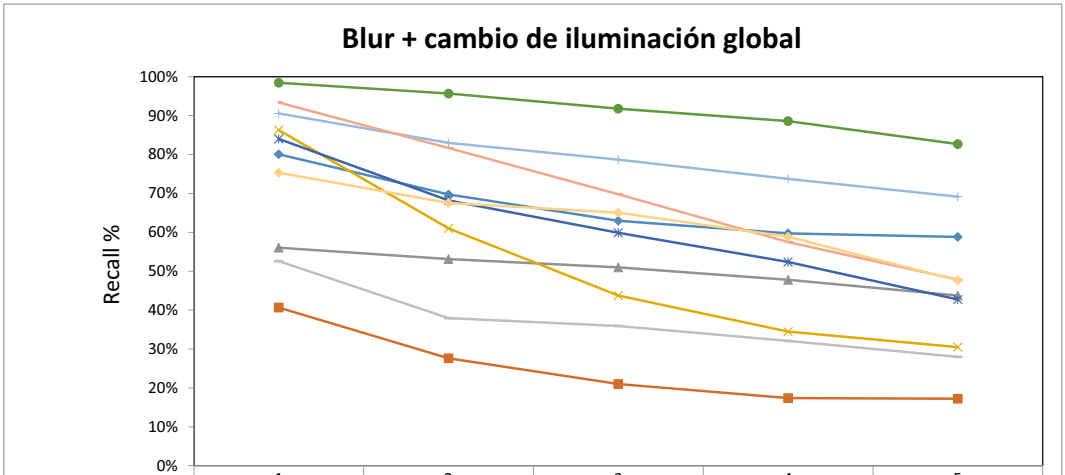
Figura 4.15: Recall secuencia blur

Las características comentadas anteriormente tienen su consecuencia en los resultados en que como se puede ver, los detectores de esquinas (FAST, Harris y sus evoluciones) son los que presentan un recall con una pendiente más acusada debido a que este tipo de estructuras se ven más afectadas por el *blurring*, mientras que los

detectores basados en el hessiano no presentan esta tendencia.

4.3.2.2. Blur combinado con cambio de iluminación global

La secuencia parte de una escena oscurecida que se va aclarando y emborronando en las sucesivas imágenes.



	1	2	3	4	5
AGAST	80,05%	69,73%	62,99%	59,71%	58,84%
BRISK	40,65%	27,63%	21,00%	17,40%	17,21%
D1	56,02%	53,10%	50,96%	47,79%	43,76%
FAST	86,32%	60,99%	43,73%	34,47%	30,49%
HarrisLaplace	83,99%	68,20%	59,87%	52,37%	42,74%
HessianLaplace	98,46%	95,65%	91,81%	88,61%	82,70%
MultiscaleHarris	90,58%	83,02%	78,66%	73,70%	69,16%
ORB	93,41%	81,68%	69,79%	57,46%	47,87%
SIFT	52,59%	37,96%	35,91%	32,06%	27,96%
SURF	75,31%	67,53%	65,06%	58,89%	47,65%

Figura 4.16: Recall secuencia blur con cambio de iluminación global

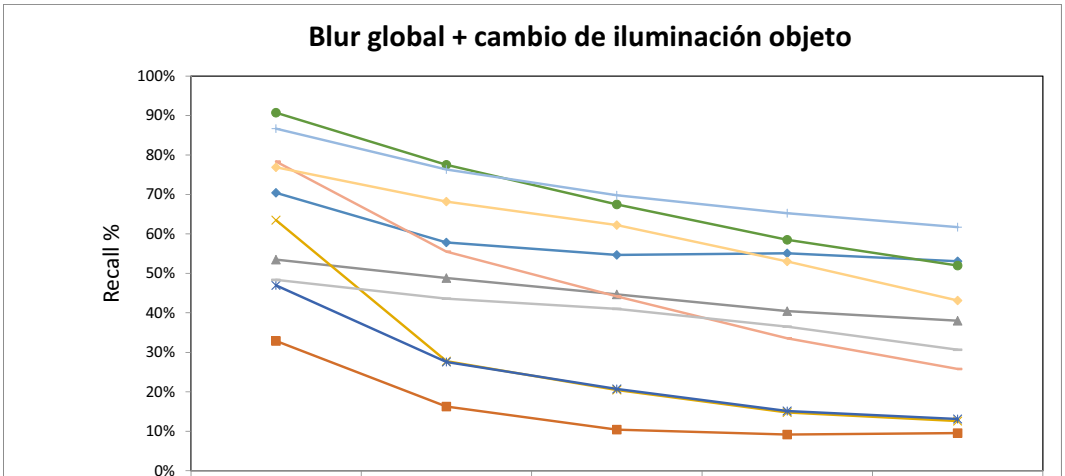
De nuevo, al igual que en la secuencia anterior se observa como los detectores de esquinas presentan una pendiente más acusada, debido a la predominancia de los efectos de *blurring* que se van endureciendo, mientras que el oscurecido se va reduciendo.

4.3.2.3. Blur global con cambio de iluminación sobre un elemento

La secuencia parte de una escena con un objeto oscurecido que se va aclarando en las siguientes escenas, mientras que se va incrementando el blurring por toda la escena.

Estas características comentadas hacen que la tendencia sea similar a las anteriores secuencias que también presentan efectos de blurring. Los algoritmos que presentan

una respuesta más estable a los cambios producidos en esta secuencia son D1 y SIFT, que tienen en común que ambos algoritmos basados en la matriz hessiana y el espacio-escala.



	1	2	3	4	5
AGAST	70,41%	57,82%	54,67%	55,11%	53,08%
BRISK	32,88%	16,27%	10,39%	9,17%	9,50%
D1	53,48%	48,82%	44,69%	40,44%	38,01%
FAST	63,47%	27,76%	20,46%	14,73%	12,59%
HarrisLaplace	46,95%	27,55%	20,73%	15,09%	13,12%
HessianLaplace	90,71%	77,52%	67,44%	58,49%	51,99%
MultiscaleHarris	86,66%	76,33%	69,82%	65,26%	61,70%
ORB	78,36%	55,49%	44,14%	33,57%	25,77%
SIFT	48,38%	43,61%	41,02%	36,50%	30,66%
SURF	76,88%	68,20%	62,23%	53,02%	43,14%

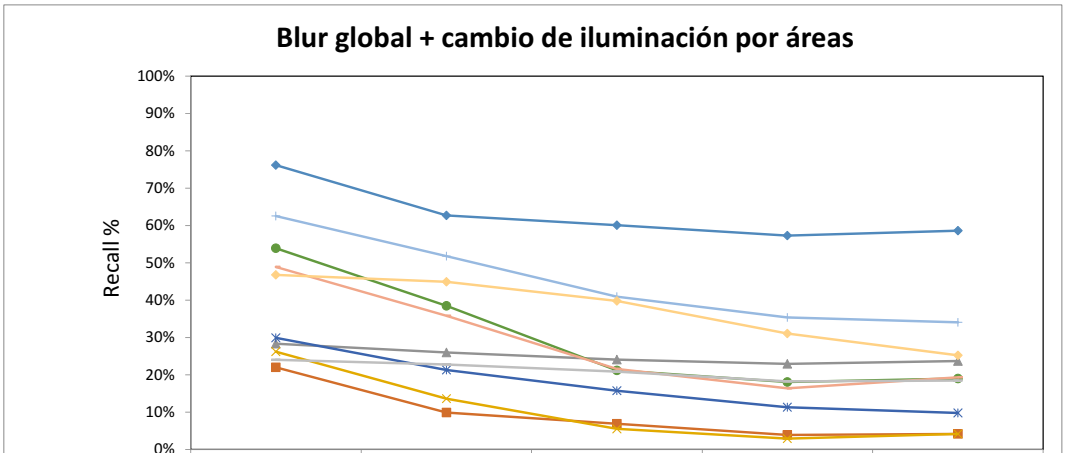
Figura 4.17: Recall secuencia blur con cambio de iluminación sobre un objeto

4.3.2.4. Blur global combinado con cambio de iluminación por áreas

La secuencia parte de una escena ensombrecida que en la que para las siguientes imágenes la sombra va retrocediendo (se reduce el área sombreada) y a su vez el blurring va en aumento. El efecto de sombra se capturó en condiciones reales mientras que el blur se aplica mediante un promediado gaussiano.

Una vez más se observa una tendencia descendente para la mayoría de algoritmos, más acusada que en otras secuencias debido a la mayor variabilidad de cambios de iluminación que resultan de capturar las imágenes en condiciones reales.

Los dos algoritmos que presentan una mayor estabilidad a las transformaciones son SIFT y D1 de nuevo.

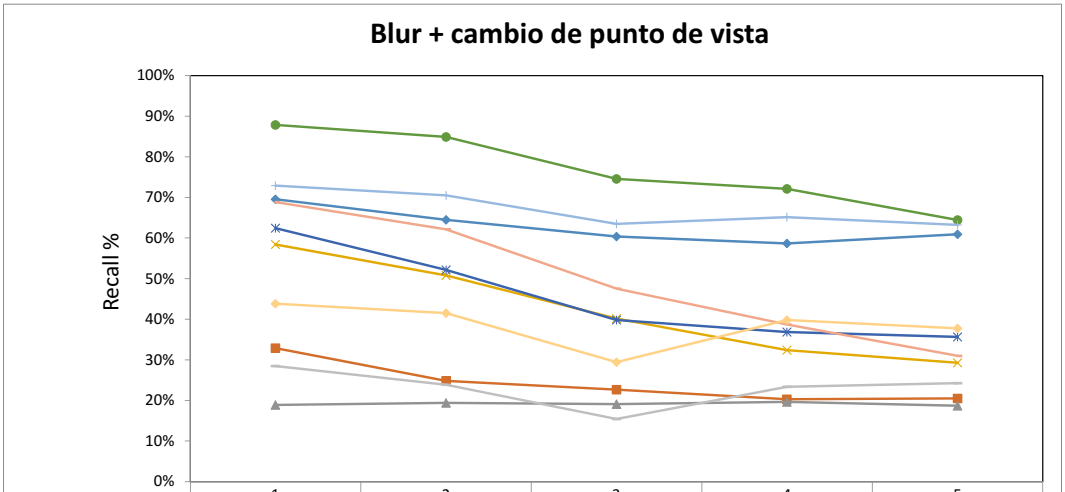


	1	2	3	4	5
AGAST	76,18%	62,68%	60,07%	57,29%	58,60%
BRISK	21,99%	9,92%	6,88%	3,90%	4,19%
D1	28,33%	26,00%	24,10%	22,91%	23,68%
FAST	26,18%	13,62%	5,51%	2,90%	4,17%
HarrisLaplace	29,94%	21,29%	15,73%	11,29%	9,79%
HessianLaplace	53,95%	38,47%	21,18%	18,07%	19,03%
MultiscaleHarris	62,52%	51,79%	40,91%	35,35%	34,07%
ORB	48,89%	35,80%	21,56%	16,36%	19,31%
SIFT	24,00%	22,76%	20,84%	18,26%	18,48%
SURF	46,78%	44,95%	39,77%	31,07%	25,24%

Figura 4.18: Recall secuencia blur global con cambio de iluminación por áreas

4.3.2.5. Blur combinado con cambio de punto de vista

Los resultados de recall de esta secuencia una vez más muestran la tendencia descendente en los resultados de recall con un pico de algunos algoritmos en el paso 3 de afectación. Este pico se deberá por tanto al efecto de la transformación de cambio de punto de vista ya que el blur como se ha visto en otras secuencias produce un decaimiento progresivo, por lo que este aspecto lo analizaremos en la secuencia de cambio de vista puro.



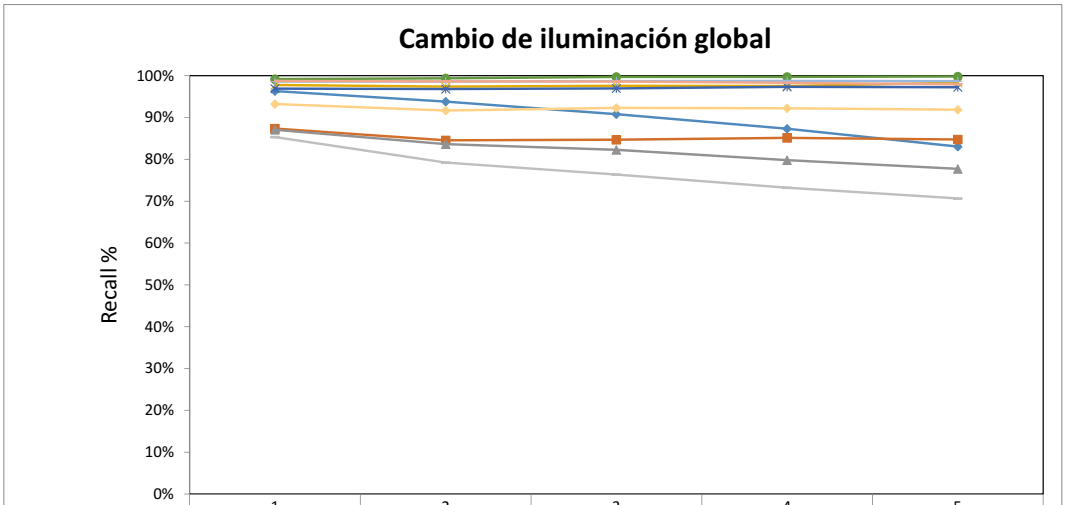
	1	2	3	4	5
AGAST	69,56%	64,47%	60,37%	58,65%	60,92%
BRISK	32,90%	24,83%	22,65%	20,31%	20,50%
D1	18,88%	19,39%	19,07%	19,65%	18,67%
FAST	58,40%	50,80%	40,18%	32,36%	29,26%
HarrisLaplace	62,41%	52,16%	39,81%	36,87%	35,65%
HessianLaplace	87,86%	84,88%	74,56%	72,09%	64,43%
MultiscaleHarris	72,90%	70,51%	63,47%	65,12%	63,25%
ORB	68,83%	62,14%	47,52%	38,70%	30,99%
SIFT	28,47%	23,86%	15,44%	23,37%	24,27%
SURF	43,80%	41,52%	29,45%	39,83%	37,77%

Figura 4.19: Recall secuencia blur con cambio de punto de vista

4.3.2.6. Cambio de iluminación global

Esta secuencia reproduce un oscurecido uniforme y progresivo por toda la imagen.

Como se puede ver, todos los algoritmos obtienen muy buenos valores de recall en comparación con otras transformaciones. Será necesario endurecer más las condiciones de cambio de iluminación para poder evaluar mejor el comportamiento de los algoritmos en estas situaciones. Este endurecimiento se puede conseguir aumentando el nivel de oscurecido uniforme o como se verá en la secuencia de ensombrecido, 4.3.2.10, mediante la captura de las imágenes en condiciones reales, que produce unos cambios de iluminación mucho más irregulares y supone un mayor reto a los detectores.

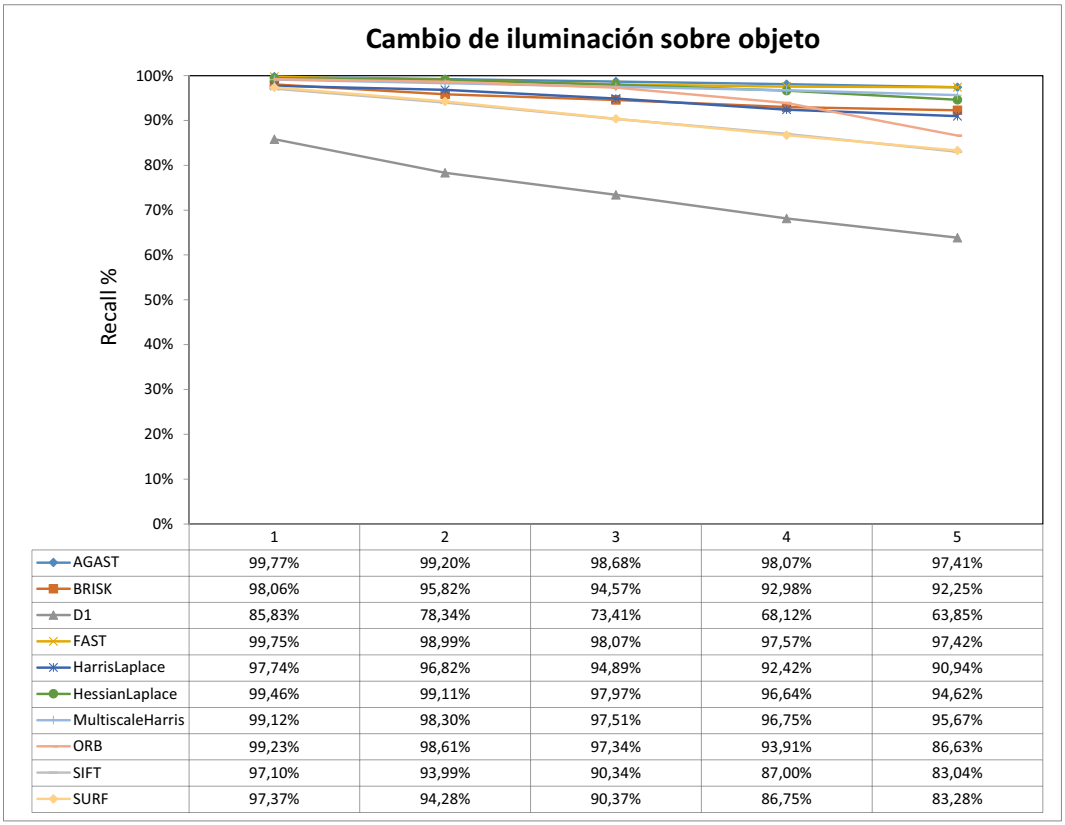


	1	2	3	4	5
AGAST	96,34%	93,83%	90,80%	87,33%	83,04%
BRISK	87,36%	84,55%	84,73%	85,14%	84,78%
D1	87,10%	83,62%	82,30%	79,83%	77,75%
FAST	97,78%	97,37%	97,58%	97,56%	98,33%
HarrisLaplace	96,92%	96,80%	96,94%	97,33%	97,25%
HessianLaplace	99,17%	99,38%	99,73%	99,71%	99,79%
MultiscaleHarris	98,61%	98,61%	98,71%	98,73%	98,70%
ORB	98,73%	98,67%	98,62%	98,31%	97,93%
SIFT	85,36%	79,22%	76,37%	73,25%	70,66%
SURF	93,22%	91,70%	92,32%	92,23%	91,86%

Figura 4.20: Recall secuencia cambio de iluminación global

4.3.2.7. Cambio de iluminación sobre objeto

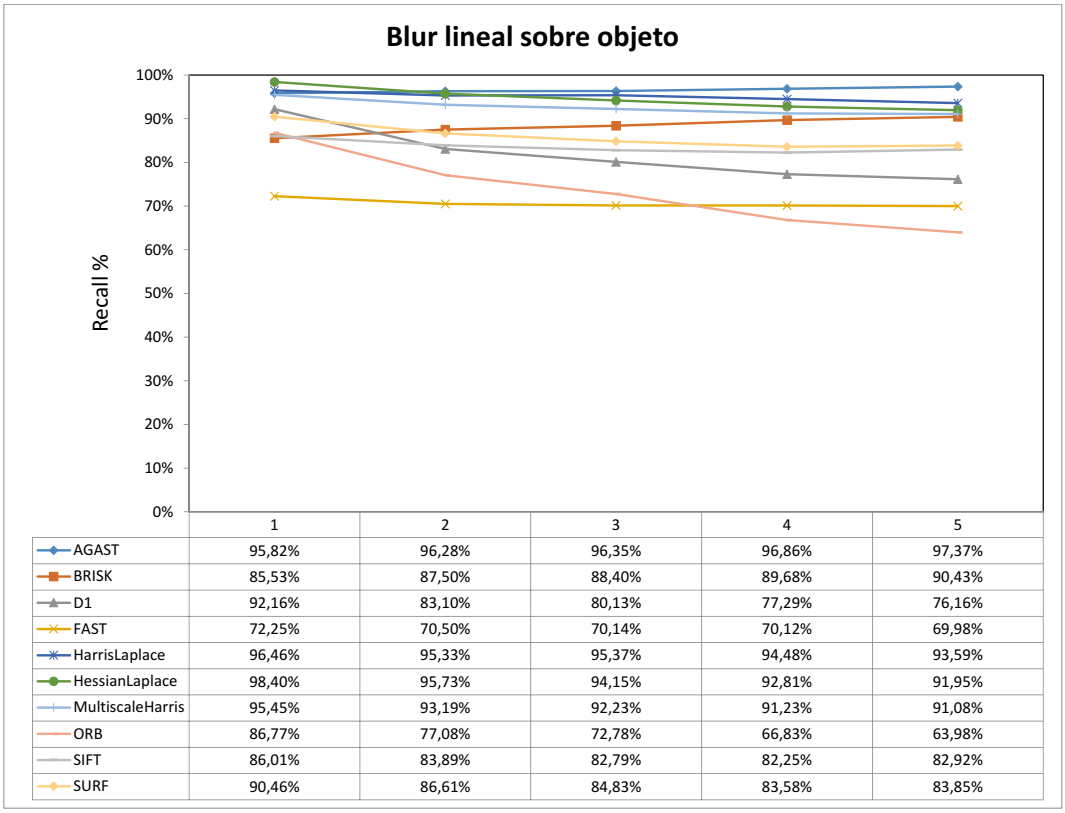
Igualmente que en la secuencia anterior, todos los algoritmos consiguen muy buenos resultados, pero en esta ocasión si que se puede observar un menor recall en las últimas secuencias de algoritmos basados en el hessiano y en el espacio escala (D1, SIFT y SURF) debido a que el objeto oscurecido presenta una textura predominantemente heterogénea, por lo que tendrán más penalización los algoritmos que detectan blobs.



	1	2	3	4	5
AGAST	99,77%	99,20%	98,68%	98,07%	97,41%
BRISK	98,06%	95,82%	94,57%	92,98%	92,25%
D1	85,83%	78,34%	73,41%	68,12%	63,85%
FAST	99,75%	98,99%	98,07%	97,57%	97,42%
HarrisLaplace	97,74%	96,82%	94,89%	92,42%	90,94%
HessianLaplace	99,46%	99,11%	97,97%	96,64%	94,62%
MultiscaleHarris	99,12%	98,30%	97,51%	96,75%	95,67%
ORB	99,23%	98,61%	97,34%	93,91%	86,63%
SIFT	97,10%	93,99%	90,34%	87,00%	83,04%
SURF	97,37%	94,28%	90,37%	86,75%	83,28%

Figura 4.21: Recall secuencia cambio de iluminación sobre objeto

4.3.2.8. Blur lineal sobre objeto



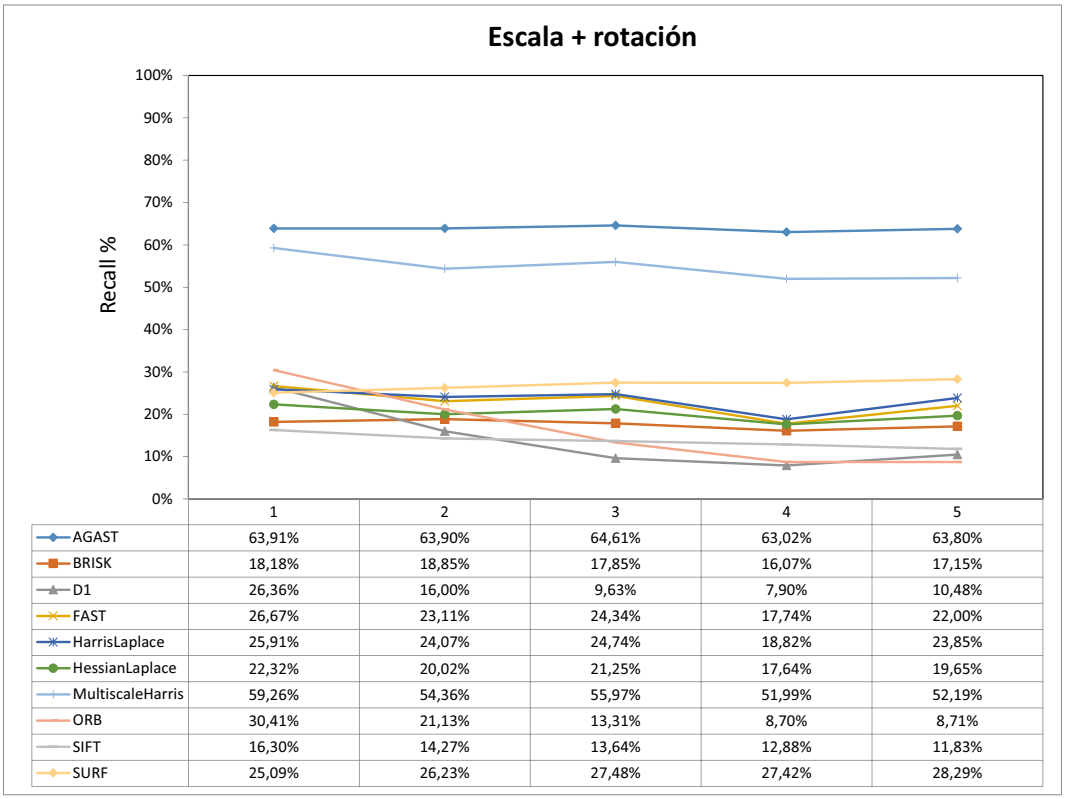
	1	2	3	4	5
AGAST	95,82%	96,28%	96,35%	96,86%	97,37%
BRISK	85,53%	87,50%	88,40%	89,68%	90,43%
D1	92,16%	83,10%	80,13%	77,29%	76,16%
FAST	72,25%	70,50%	70,14%	70,12%	69,98%
HarrisLaplace	96,46%	95,33%	95,37%	94,48%	93,59%
HessianLaplace	98,40%	95,73%	94,15%	92,81%	91,95%
MultiscaleHarris	95,45%	93,19%	92,23%	91,23%	91,08%
ORB	86,77%	77,08%	72,78%	66,83%	63,98%
SIFT	86,01%	83,89%	82,79%	82,25%	82,92%
SURF	90,46%	86,61%	84,83%	83,58%	83,85%

Figura 4.22: Recall secuencia blur lineal sobre objeto

En este caso, todos los algoritmos obtienen unos valores muy altos de recall. Se puede ver, sin embargo, como D1 y ORB son los que presentan una pendiente descendente mayor, lo que se traduce en una mala respuesta ante el endurecimiento de las condiciones.

4.3.2.9. Escala combinado con rotación

Como se explicó en la sección3.1 los efectos de escala y rotación se producen sobre un objeto de la imagen y es sólo sobre las posiciones que ocupa el objeto sobre las que se calcula el recall. La imagen original es en la que el objeto está más alejado de tal manera que no desaparecen estructuras detectables a medida que se acerca sino al contrario.



	1	2	3	4	5
AGAST	63,91%	63,90%	64,61%	63,02%	63,80%
BRISK	18,18%	18,85%	17,85%	16,07%	17,15%
D1	26,36%	16,00%	9,63%	7,90%	10,48%
FAST	26,67%	23,11%	24,34%	17,74%	22,00%
HarrisLaplace	25,91%	24,07%	24,74%	18,82%	23,85%
HessianLaplace	22,32%	20,02%	21,25%	17,64%	19,65%
MultiscaleHarris	59,26%	54,36%	55,97%	51,99%	52,19%
ORB	30,41%	21,13%	13,31%	8,70%	8,71%
SIFT	16,30%	14,27%	13,64%	12,88%	11,83%
SURF	25,09%	26,23%	27,48%	27,42%	28,29%

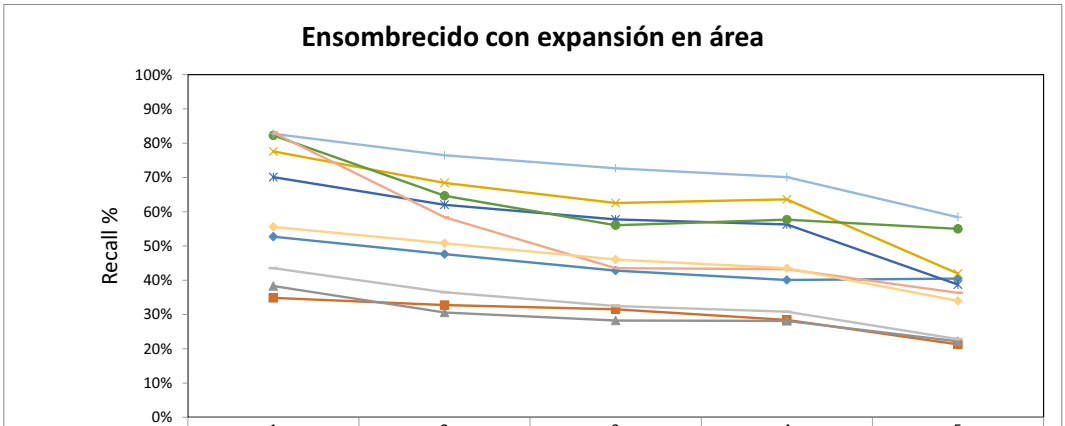
Figura 4.23: Recall secuencia escala+rotación

Sobresalen sobre los demás los resultados de AGAST y Multi-scale Harris, ambos detectores de esquinas, lo que se deberá a las características de la imagen que hace que estos dos algoritmos detecten una gran cantidad de este tipo de estructuras en todas las imágenes.

El resto de algoritmos presentan recalls menores pero igualmente estables por lo que se puede decir que tienen una respuesta estable al endurecimiento de las condiciones.

4.3.2.10. Ensombrecido con expansión en área

Como se ha comentado en anteriores secciones, el cambio de iluminación de esta secuencia se capturó en condiciones reales.



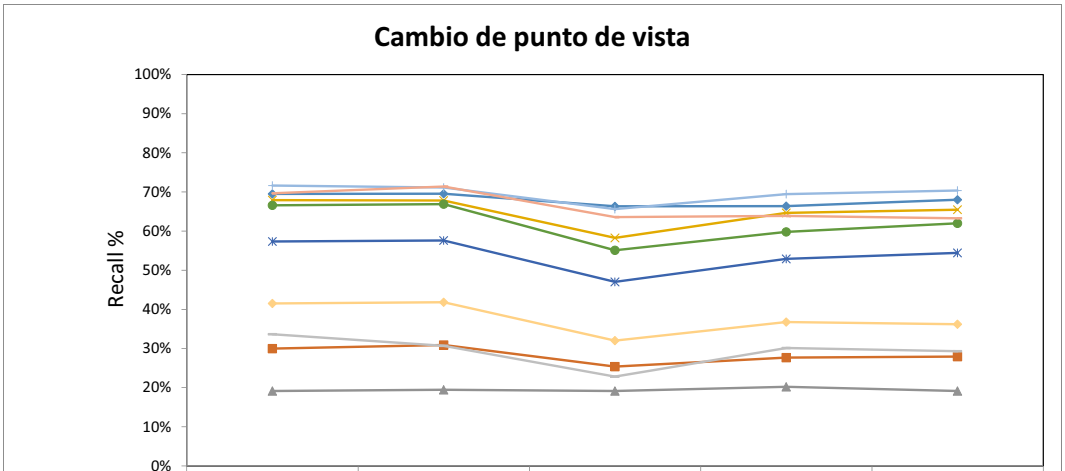
	1	2	3	4	5
AGAST	52,73%	47,65%	42,81%	40,04%	40,50%
BRISK	34,87%	32,77%	31,51%	28,40%	21,27%
D1	38,30%	30,55%	28,22%	28,14%	22,13%
FAST	77,60%	68,40%	62,55%	63,60%	41,92%
HarrisLaplace	70,13%	62,01%	57,77%	56,30%	38,72%
HessianLaplace	82,25%	64,69%	56,07%	57,72%	55,01%
MultiscaleHarris	82,75%	76,49%	72,66%	70,08%	58,39%
ORB	83,10%	58,42%	43,52%	43,23%	36,33%
SIFT	43,55%	36,51%	32,47%	30,79%	22,78%
SURF	55,57%	50,76%	46,01%	43,53%	34,01%

Figura 4.24: Recall secuencia ensombrecido con expansión en área

Por la toma de las imágenes en condiciones reales se producen unos cambios de iluminación mucho más irregulares a lo largo de la imagen, que a su vez supone un reto mayor para los algoritmos que si el oscurecido se produce uniformemente (cuando es modelado en ordenador o se varía la apertura de la cámara)

4.3.2.11. Cambio de punto de vista

Como se explicó en la sección 3.1, la secuencia modela un cambio de punto de vista de aproximadamente 70 grados que empieza y termina en posiciones laterales.



	1	2	3	4	5
AGAST	69,51%	69,53%	66,32%	66,37%	68,00%
BRISK	29,98%	30,87%	25,38%	27,66%	27,91%
D1	19,14%	19,42%	19,16%	20,23%	19,13%
FAST	67,93%	67,82%	58,27%	64,66%	65,46%
HarrisLaplace	57,34%	57,61%	47,05%	52,88%	54,43%
HessianLaplace	66,58%	66,89%	55,11%	59,82%	61,99%
MultiscaleHarris	71,66%	71,15%	65,62%	69,46%	70,38%
ORB	69,65%	71,37%	63,57%	63,90%	63,28%
SIFT	33,67%	30,69%	22,83%	30,13%	29,29%
SURF	41,51%	41,81%	32,02%	36,78%	36,17%

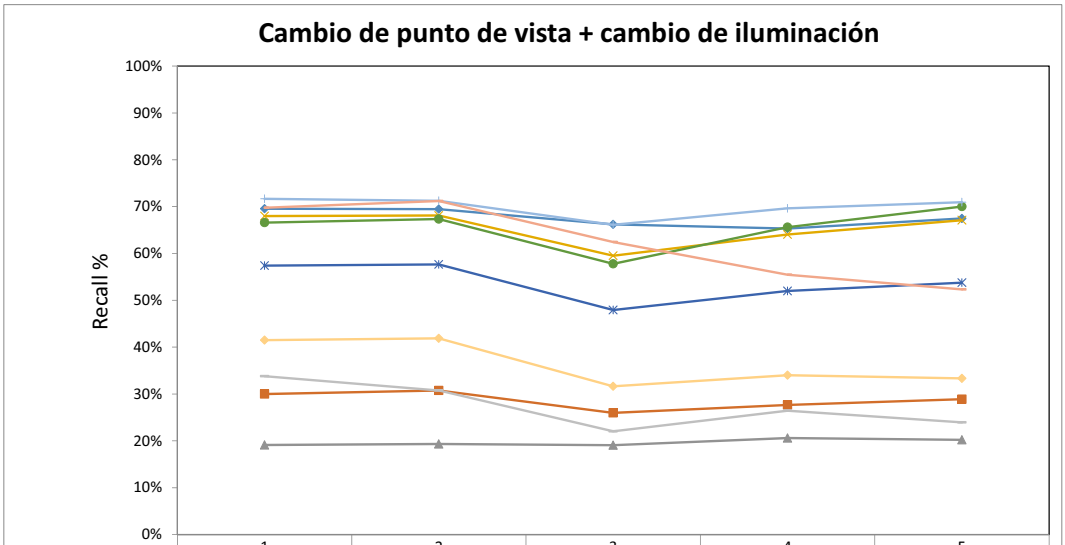
Figura 4.25: Recall secuencia cambio de punto de vista

Se observa descenso del recall en casi todos los algoritmos para el tercer paso de camino de punto de vista, esto puede ser debido a que es la imagen más frontal de la secuencia y por tanto en la que más puntos se detectan que no tienen correspondencia en la original.

Los resultados de recall para todos los algoritmos se mantienen bastante estables a lo largo de a secuencia, por lo que se podrían endurecer más las condiciones de cambio de punto de vista para observar otras tendencias.

4.3.2.12. Cambio de punto de vista combinado con cambio de iluminación

Se trata de una secuencia análoga a la anterior pero en la que las imágenes se van oscureciendo en cada paso de cambio de punto de vista.



	1	2	3	4	5
AGAST	69,54%	69,45%	66,19%	65,33%	67,46%
BRISK	29,99%	30,76%	25,97%	27,66%	28,90%
D1	19,12%	19,35%	19,08%	20,62%	20,22%
FAST	67,97%	68,13%	59,51%	64,06%	67,10%
HarrisLaplace	57,43%	57,66%	47,93%	51,99%	53,76%
HessianLaplace	66,60%	67,34%	57,79%	65,63%	70,02%
MultiscaleHarris	71,64%	71,23%	66,12%	69,64%	70,93%
ORB	69,75%	71,21%	62,45%	55,44%	52,35%
SIFT	33,80%	30,74%	22,05%	26,42%	23,96%
SURF	41,51%	41,87%	31,63%	34,02%	33,33%

Figura 4.26: Recall secuencia cambio de punto de vista + cambio de iluminación

Los resultados de recall son ligeramente inferiores a los de la secuencia de cambio de punto de vista puro pero muy similares salvo para el caso de ORB en el que si que se aprecia un descenso significativo respecto a aquella en los dos últimos pasos.

4.3.2.13. Conclusiones generales.

Como se apuntó en la sección 4.3.1.9 el dataset aportado para este proyecto se ha construido evitando movimientos de cámara cuando la transformación no lo requiere. Gracias a esto se evita el paso de obtención de homografías en estas secuencias, lo que evita posibles fallos y además permite que en la evaluación se aíslen los efectos que se desea medir de otros efectos derivados de los movimientos de cámara.

Además se han evaluado los algoritmos sobre una variedad de nuevas transformaciones, aumentando la complejidad de las mismas y añadiendo nuevas combinaciones, lo que ha aportado una mayor riqueza a la evaluación al poder extraer nuevas conclusiones en base a los comportamientos de los algoritmos en estas situaciones.

4.4. Conclusiones.

A lo largo de la evaluación sobre los dos conjuntos de datos se ha ido viendo como algunos algoritmos típicamente obtienen en todas las secuencias unos valores de recall altos (Multi-scale Hessian, Hessian-Laplace, AGAST, Multi-scale-Harris) mientras que otros típicamente obtienen valores más bajos (BRISK, D1, SIFT). Esto se debe a que, por la forma en que están configurados los algoritmos, los primeros hacen una detección masiva de puntos que se suele traducir en valores altos de recall, mientras que los segundos obtienen menor número de puntos de interés pero es más común observar como los valores de recall no decaen con tanta velocidad en los sucesivos pasos de las secuencias como sí ocurre en los otros casos, por lo que podría sugerir una mayor robustez de las detecciones correctas.

También se ha visto que el tipo de estructuras que detectan los algoritmos (blobs o esquinas y bordes) tienen una influencia importante en los resultados cuando en las imágenes hay una predominancia clara de alguna de dichas estructuras. Además se ha observado que algunos detectores de esquinas como FAST, ORB y Harris-Laplace se ven más afectados (pendientes mayores) ante efectos de blurring.

El algoritmo AGAST, en general ha obtenido los mejores resultados de entre todos los detectores, mientras que BRISK, cuyo detector está basado en el propio AGAST, ha sido de los peores. Esto puede ser debido al sesgo de puntos por las modificaciones que implementa el detector de BRISK de cara a la posterior fase de descripción.

De entre los detectores que forman parte de algoritmos que constan de detector+descriptor como son BRISK, SIFT, SURF y ORB, los que han obtenido en general unos valores más altos de recall son SURF y ORB, pero en el caso de ORB resalta una peor invariancia (mayor pendiente) ante cambios de iluminación, por lo que será el detector de SURF el que se utiliza como estándar para evaluar descriptores en el capítulo siguiente. No obstante, en los cuatro casos se trata de unos detectores bastante complejos y es de suponer que todos ellos estén optimizados de cara a proveer unos puntos adecuados para su descriptor correspondiente.

Si se comparan los resultados con la evaluación teórica, donde se hizo una clasificación de los algoritmos en función de si implementaban estrategias para ser más invariantes a rotación, escala y cambios de punto de vista, se pueden extraer varias conclusiones.

Los algoritmos que constan de fase de detección y descripción, salvo SURF, en general obtienen peores resultados que otros detectores ante éstas transformaciones, pero como se ha comentado, es de esperar que las estrategias que implementan puedan tener una mayor influencia si se evalúan los algoritmos completos. En cuanto a SURF,

se ha visto que responde bien ante estas transformaciones, y que, ya desde esta fase de detección, muestra los resultados en invariancia prometidos en la evaluación teórica.

En cuanto a los algoritmos de detección que no tienen un descriptor asociado, AGAST ha mostrado el mejor resultado para las tres invariancias mencionadas, mientras que en la evaluación teórica sólo se resaltó la invariancia a rotación. Se puede concluir que partiendo de las ideas de FAST, las mejoras que implementan demuestran buenos resultados como se anuncia en la publicación de los autores. Esto hará también que el detector de BRISK, aunque en esta etapa de detección no resalten sus resultados, parta de una buena base de cara a la descripción.

Se ha visto también cómo el tipo de operador asociado a la detección tiene una gran influencia en los resultados mostrados por los algoritmos en estas situaciones. Si se atiende a las estrategias que implementan los algoritmos para hacer frente a éstas transformaciones, se puede ver como algunos de ellos parten con ventaja para conseguir estas invariancias gracias a que el tipo de detección de base es mejor para hacer frente a estas transformaciones.

Capítulo 5

Descriptores.

En el Capítulo 2, la etapa de descripción se define como la caracterización de las detecciones mediante un vector de características denominado descriptor. Las técnicas de descripción persiguen dos objetivos principales, la discriminatividad y la repetibilidad. Para ello, las estrategias tradicionales trabajaban con complejas técnicas haciendo uso de información de gradientes o similares. En la actualidad comienza a trabajarse también con técnicas de menor coste como los descriptores binarios. Con el elevado número de algoritmos propuestos en este área, al igual que ocurre con las técnicas de detección, resulta complejo discernir cuales son las que presentan mejores comportamientos para los objetivos deseados.

Con el objetivo de facilitar esta tarea, en este capítulo se va llevar a cabo una categorización de las técnicas del estado del arte (sección 5.1) que permitirá a su vez realizar una pre-selección de los algoritmos más relevantes dentro de cada categoría. Los algoritmos seleccionados serán analizados teóricamente en la misma sección. Dicho estudio teórico se complementará con un análisis comparativo desde el punto de vista teórico (sección 5.2), donde se ha desarrollado una evaluación teórica de cada una de las técnicas escogidas frente a las distintas propiedades de interés de los detectores. Los algoritmos seleccionados serán evaluados en el marco de evaluación propuesto (sección 5.3). El capítulo se cierra con unas conclusiones (sección 5.4) sobre los resultados obtenidos.

5.1. Categorización y selección de técnicas.

Para el desarrollo de este capítulo, se han agrupado los descriptores en tres categorías. En primer lugar, como técnica central del estado del arte, se encuentra SIFT. Su publicación supuso un punto de inflexión en la detección y descripción de pun-

tos de interés y durante años ha sido la técnica más utilizada en gran variedad de aplicaciones.

En la segunda categoría se agrupan una serie de técnicas que, basándose en SIFT, han aportado mejoras y evoluciones a la descripción de puntos de interés. El primero de ellos fue SURF, que surgió como una primera solución al coste computacional de SIFT, consiguiendo además mejores resultados en determinadas situaciones. Como segunda técnica de ésta categoría, se evaluará el algoritmo DAISY[6], basado en las fortalezas de GLOH (que está basado en SIFT) y del propio SIFT, y que pretende mejorar los resultados de éste en invariancias, robustez a oclusiones y cambios de punto de vista, y además consiguiendo mejorar notablemente el coste computacional de sus precursores.

Por último se evaluarán, dentro de una tercera categoría de descriptores binarios, los métodos BRISK y FREAK. Mientras que los anteriores algoritmos proponen crear un descriptor sobre la región de interés de la imagen, estos métodos proponen trabajar directamente con una descripción binaria de la imagen. Con ello, e implementando diversas estrategias en cada caso, consiguen en mayor o menor medida buenos resultados a un muy bajo coste computacional.

5.1.1. SIFT

En esta fase, el algoritmo de detección y descripción SIFT [2], los puntos de interés ya han sido localizados en el espacio escala, por lo que cada punto ya está caracterizado por su localización espacial y escala $L(x, y, \sigma)$. El siguiente paso para llegar a la descripción es el cálculo de las magnitudes $m(x, y)$, y orientaciones $\theta(x, y)$ del gradiente de cada uno de los puntos detectados:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = arctan\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}$$

A continuación, se crea un histograma de orientaciones que cubre las direcciones y magnitudes del gradiente por sectores, cubriendo 360 grados alrededor de cada punto, y se selecciona la dirección o direcciones dominantes del punto. En la última etapa, análogamente, se realiza un muestreo de las orientaciones y magnitudes del gradiente de la imagen sobre regiones alrededor del punto de interés y se analizan para formar histogramas de orientaciones de cada sub-región, ver Figura 5.1.

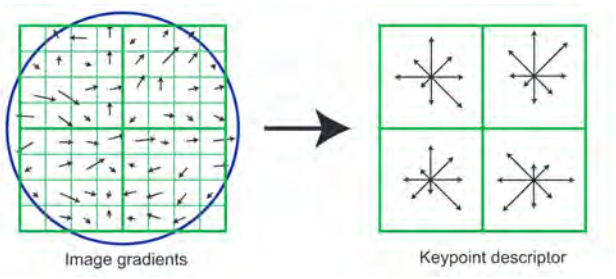


Figura 5.1: SIFT Descriptor. (a) Histogramas de orientación, (b) gradientes orientados. Fuente [2]

El resultado final es un descriptor de cada punto de interés que está formado por un vector de características que contiene los valores de todas las orientaciones de cada uno de los histogramas. Además se lleva a cabo una normalización de este vector como estrategia para dotar al descriptor de robustez frente a cambios de iluminación que puedan afectar a la magnitud de los gradientes.

5.1.2. Descriptores SIFT-based

5.1.2.1. SURF

Al igual que el detector, el descriptor de SURF hace uso de imágenes integrales, que permitirán calcular respuestas a filtros tipo Haar Wavelet5.2 rápidamente, que se usarán como parte del proceso para conseguir invariancia a rotación. El algoritmo calcula y pondera la respuesta a estos filtros en un vecindario del punto de interés de radio $6s$ (donde s es la escala característica del punto).

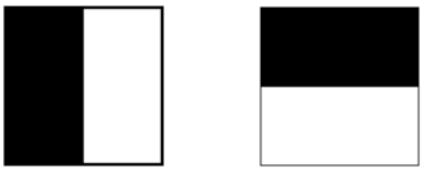


Figura 5.2: Filtros Haar Wavelet para calcular las respuestas en las direcciones x (izquierda) e y (derecha). Las partes oscuras tienen peso -1 y las claras +1. Fuente [1]

Éstas respuestas son representadas como puntos en el espacio con la respuesta horizontal en el eje de abscisas y la vertical en el de ordenadas, para después calcular la orientación principal mediante una ventana deslizante de amplitud $\frac{\pi}{3}$ (ver figura5.3).

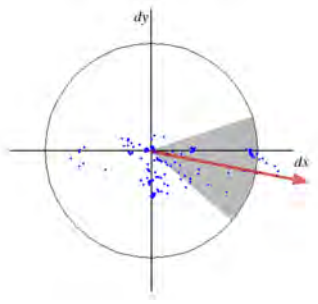


Figura 5.3: Asignación de la orientación en base a las respuestas a los filtros *wavelet*. Fuente[1]

Para construir el descriptor también se hace uso de las respuestas a los filtros *wavelet*, que se calculan en un en un vecindario alrededor del punto con un área de tamaño$20s$ (donde s es la escala) dividida en 16 regiones (4x4). A su vez, para conseguir invariancia a cambios de iluminación, cada una de las 16 regiones se dividen en 4 sub-regiones cada una (2x2) que se corresponden con un vector v de los sumatorios de las respuestas a los filtros. Además las orientaciones se normalizan para ser invariantes también a cambios de escala y contraste.

$$v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|).$$

.

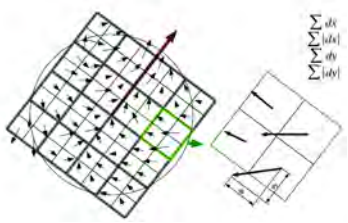


Figura 5.4: Descriptor de SURF. Fuente [1]

5.1.2.2. DAISY

Se trata de un descriptor basado en SIFT[2] que pretende ser computacionalmente más eficiente. Para ello proponen calcular los mapas de orientación de los píxeles mediante convoluciones con filtros gaussianos en áreas concéntricas, como en la figura 5.5, donde la cantidad de suavizado gaussiano es proporcional al radio del círculo. Ésta distribución circular y el uso de kernels gaussianos isotrópicos pretenden conseguir descriptores invariantes a rotación. Además proponen el uso de máscaras sobre dichos descriptores con el objetivo de hacerlos robustos frente a oclusiones.

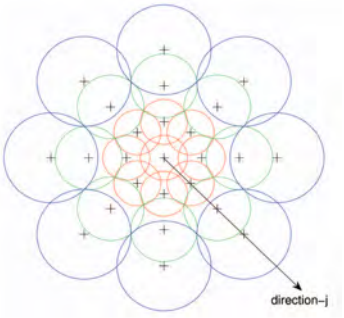


Figura 5.5: Regiones para las que el descriptor DAISY calcula la convolución con filtros gaussianos proporcionales al radio. Fuente[6]

5.1.3. Descriptores binarios

5.1.3.1. BRISK

Se trata de un descriptor binario inspirado en el descriptor BRIEF[32], basado en tests de comparación de brillo de los píxeles. Como mejoras, pretende conseguir robustez frente a rotaciones identificando la orientación característica de cada punto clave y maximizar la descriptividad a partir de esa orientación, mediante comparaciones de brillo en distribuciones de suavizados gaussianos similares a los de DAISY[6] (ver figura5.6).

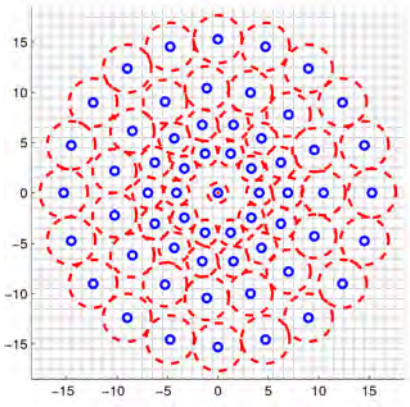


Figura 5.6: Distribución de N=60 puntos de muestreo (círculos azules) donde los la desviación estándar del kernel gaussiano que aplica el suavizado es proporcional al radio de los círculos rojos. Fuente[5]

5.1.3.2. FREAK

Se trata de un descriptor que de forma similar a BRISK utiliza distribuciones circulares concéntricas de suavizados gaussianos para efectuar comparaciones de intensidad entre pares de píxeles. Sin embargo proponen una distribución de suavizados inspirada en la retina humana, en la que dichas regiones se distribuyen con mayor densidad cerca del centro del grid, lo que produce solapamiento, y desciende exponencialmente según aumenta la distancia. El solapamiento además pretende añadir redundancia, con la que aumentar igualmente el poder discriminativo del descriptor.

Con el objetivo de estimar la rotación del punto clave utilizan una estrategia similar a BRISK, donde calculan la suma de los gradientes locales de los pares píxeles sobre los que se compara la intensidad.

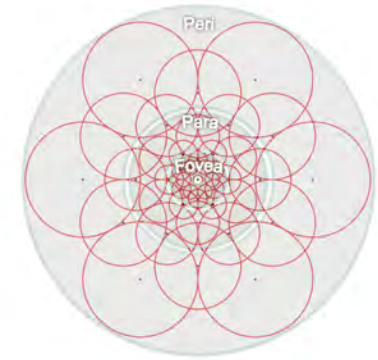


Figura 5.7: Distribución de los patrones de muestreo mediante suavizados gaussianos, inspirados en las regiones de receptores de la retina. Fuente [7]

5.2. Evaluación comparativa teórica.

Una vez analizadas todas las técnicas, al igual que se hizo en la sección 4.2, es posible llevar a cabo un análisis teórico de las propiedades que cada uno de los algoritmos presenta. Los criterios seguidos han sido los mismos que los presentados en dicha sección. Los aspectos concretos valorados adicionalmente a los expuestos entonces han sido:

- Repetibilidad teórica, haciendo referencia a las cuatro principales invariancias que puede presentar: a iluminación, a rotación, a escala, y a punto de vista (*viewpoint*) y a rotación; así como a la robustez al emborronado (*blurring*).

Descriptor	*Año*	*Tipo de descriptor*	*Detector asociado*	*Repetibilidad*					*Fortalezas*	*Precursores*	*Mejoras*
				Invarianza				*Blurring*			
				Iluminación	*Rotación*	*Escala*	*Viewpoint*				
SIFT	*2004*	*HoG*	*DoG SIFT*	*+++*	*++*	*+++*	*+*	*+++*	*Precisión*	-	-
SURF	*2006*	*HoG*	*SURF*	*+++*	*++*	*++*	*+*	*++*		*SIFT*	*Imágenes integrales + Haar Wavelet*
DAISY	*2010*	*HoG*	-	*+++*	*++*	*++*	*++*	*++*	*Localidad, precisión*	*SIFT, GLOH*	*Máscaras frente oclusiones*
BRISK	*2011*	*Binario*	*Oriented FAST*	*++*	*++*	*+++*	*+*	*+*	*Eficiencia*	*BRIEF,DAISY*	*Nuevo patrón de suavizados + orientación*
FREAK	*2012*	*Binario*	-	*+++*	*+++*	*+++*	*++*	*++*	*Eficiencia*	*BRISK*	*Nuevo patrón de menor coste*

Tabla 5.1: Análisis comparativo teórico de las técnicas de descripción. Un (+) indica baja invariancia y (+++) indica alta invariancia. De izquierda a derecha se ven los descriptores y el año en el que fueron propuestos. La técnica base para generar el descriptor, el detector asociado en caso de tenerlo, las invariancias que implementan y su robustez ante el blurring. Las últimas columnas atienden a criterios de fortalezas principal del método, precursores y qué aportaron sobre ellos.

5.3. Evaluación y análisis.

Como uno de los objetivos principales del proyecto, a continuación de la evaluación teórica presentada, se llevará a cabo una evaluación práctica sobre el nuevo marco común y mejorado aportado en este proyecto.

De manera análoga al capitulo anterior, la evaluación se realizará en primer lugar sobre el conjunto de datos presentado por K. Mikolajczyk [18] como principal referencia en el estado del arte y en segundo lugar sobre el nuevo conjunto de datos construido para el nuevo marco de evaluación de este proyecto.

Los datos se calcularán y presentarán en base a las métricas expuestas en la sección3.2.2.2, que a su vez permitirán analizar resultados y comparar con las conclusiones extraídas de la evaluación teórica.

5.3.1. Evaluación sobre el conjunto de datos de K. Mikolajczyk

5.3.1.1. Blur sobre escena con regiones homogéneas

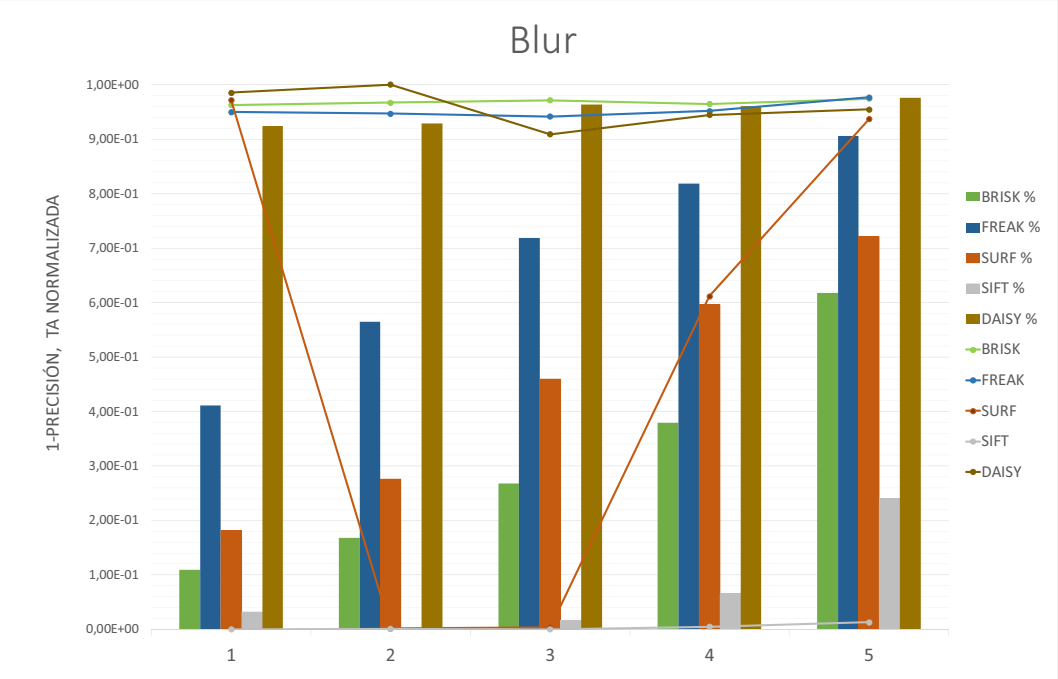


Figura 5.8: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan la tasa de asociaciones inversa y normalizada.

Como se puede observar, SIFT obtiene unos resultados de precisión y tasa de asociaciones muy superior al resto de algoritmos para esta secuencia. En segundo lugar en cuanto a tasa de asociaciones estaría BRISK, muy alejado de SIFT. Los peores resultados en ambos sentidos los obtiene DAISY.

5.3.1.2. Blur sobre escena texturada

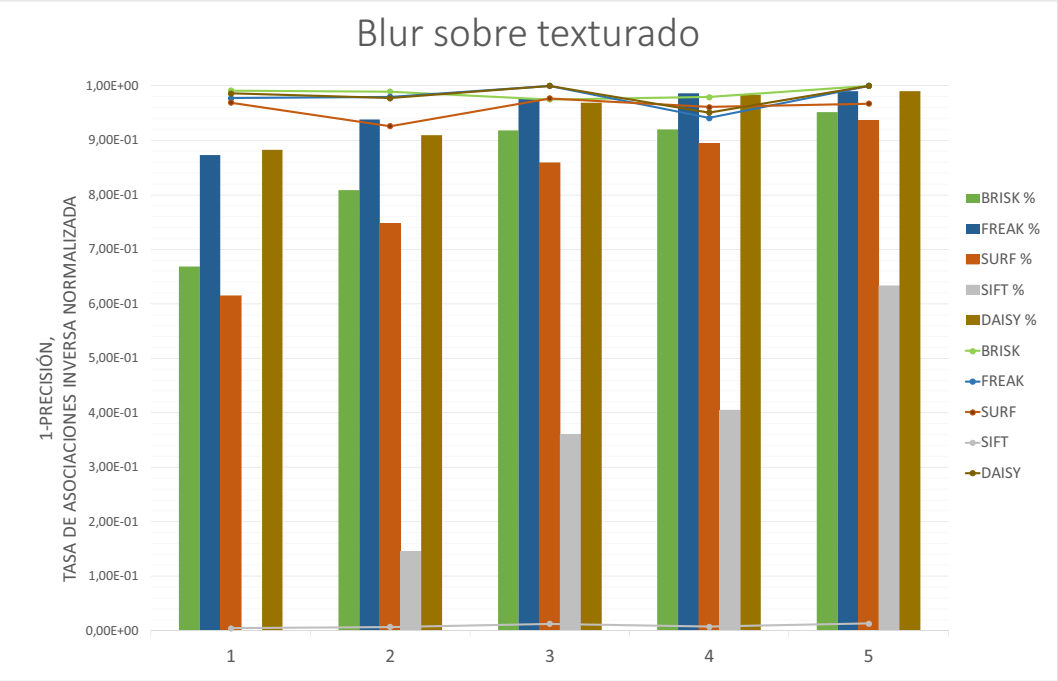


Figura 5.9: Resultados secuencia de 'Blur sobre texturado' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan la tasa de asociaciones inversa y normalizada.

En esta segunda secuencia de *blur*, los resultados para todos los algoritmos son peores que en la anterior, pero igualmente, SIFT obtiene unos resultados mucho mejores, resaltando especialmente en precisión.

5.3.1.3. Rotación combinado con zoom

Se puede ver que las tres última imágenes de secuencia modelan unas condiciones demasiado duras como para realizar una evaluación sobre ellas,debido a los deficientes resultados que muestran todos los algoritmos. En la primer nivel de degradación SIFT obtiene unos resultados mucho mejores que el resto de algoritmo, mientras que en el segundo nivel se observa un mejor funcionamiento de SURF.

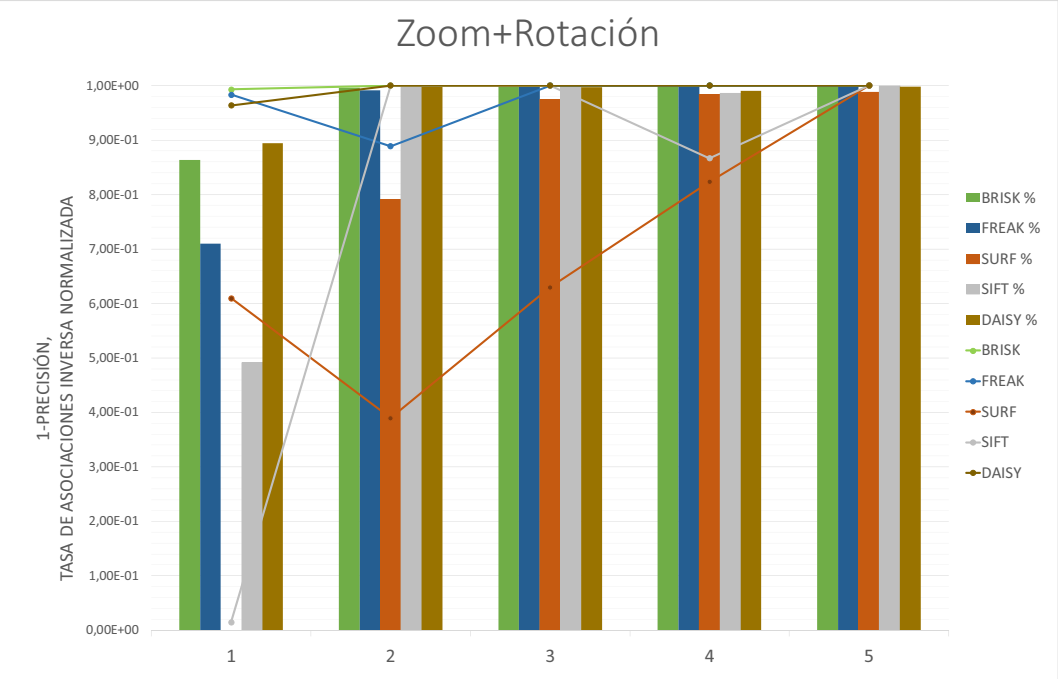


Figura 5.10: Resultados secuencia de 'Zoom+rotación' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.1.4. Rotación combinado con zoom sobre escena texturada

Esta secuencia, como se comentó en el capítulo de detección, supone unas condiciones muy duras para el funcionamiento de los algoritmos, sólo en el primer paso se puede ver como SURF obtiene una tasa de asociaciones del 100 % pero con una penalización importante de precisión.

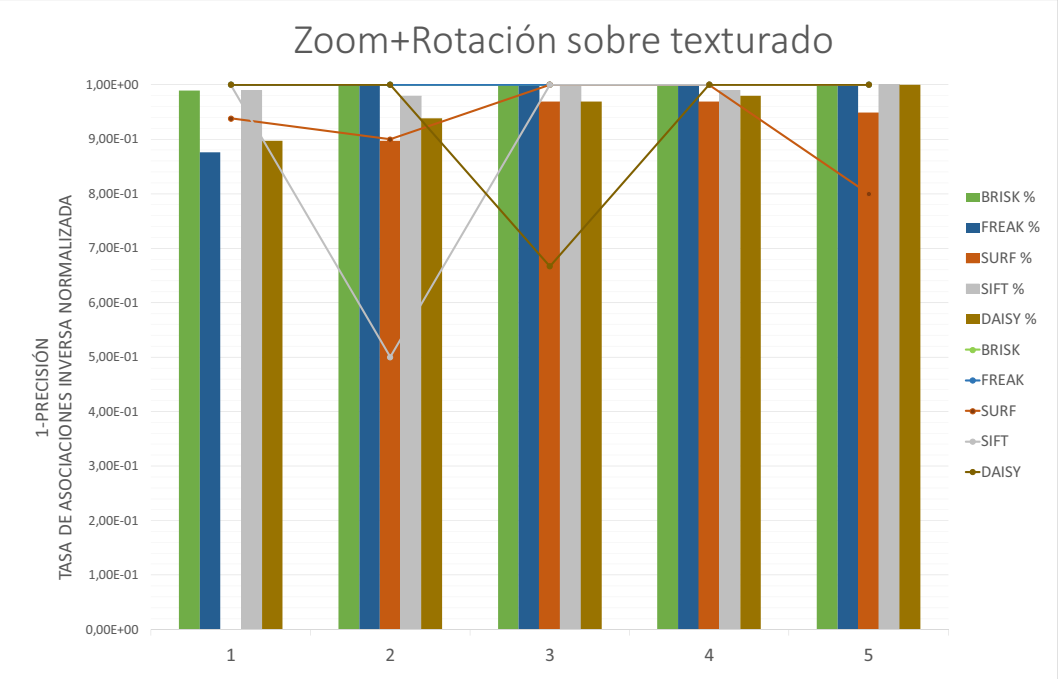


Figura 5.11: Resultados secuencia de 'Zoom+rotación sobre texturado' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.1.5. Cambio de punto de vista

De nuevo se trata de una secuencia que ofrece unas condiciones muy duras, en la que la tasa de asociaciones es muy baja para todos los algoritmos salvo para el primer nivel de la transformación que SURF obtiene una tasa del 100 %. A partir del tercer paso de cambio de punto de vista los resultados son muy deficientes en todos los algoritmos como para poder extraer conclusiones.

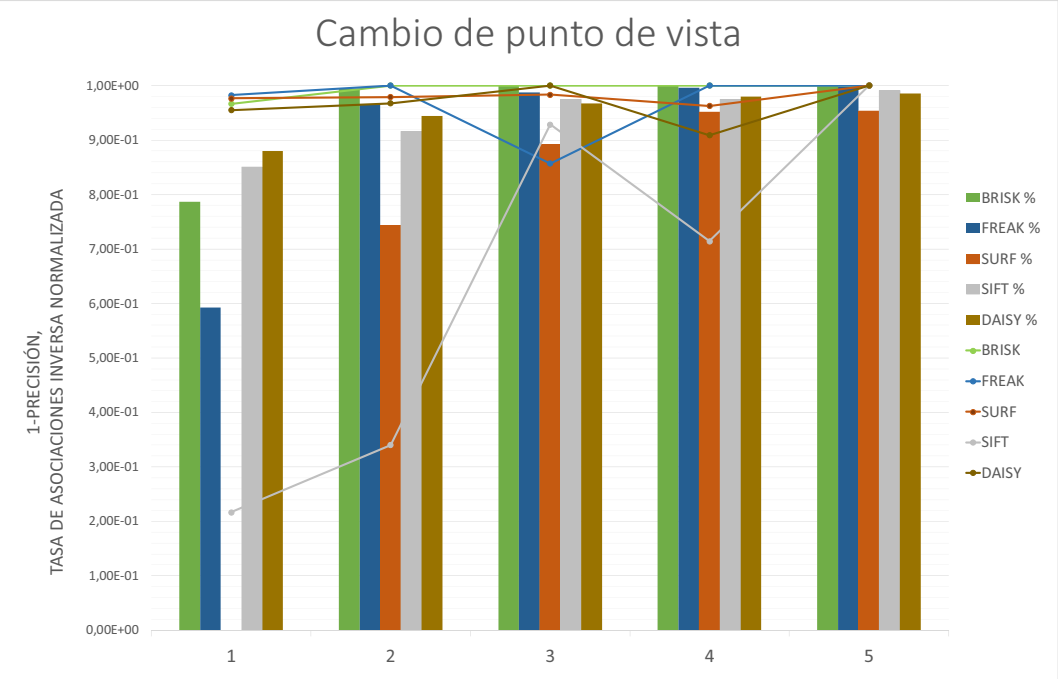


Figura 5.12: Resultados secuencia de 'Cambio de punto de vista' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/-correspondencias.

5.3.1.6. Cambio de punto de vista sobre escena texturada

Los resultados de precisión y tasa de asociaciones en esta secuencia para BRISK, FREAK y DAISY son de nuevo demasiado pobres como para observar algún tipo de evolución de la que extraer conclusiones. Por otro lado SIFT y SURF si que presentan una tendencia descendiente en los sucesivos niveles de afectación , obteniendo SIFT, en comparación con el resto de algoritmos,muy buenos resultados de nuevo para los dos primeros niveles de afectación.

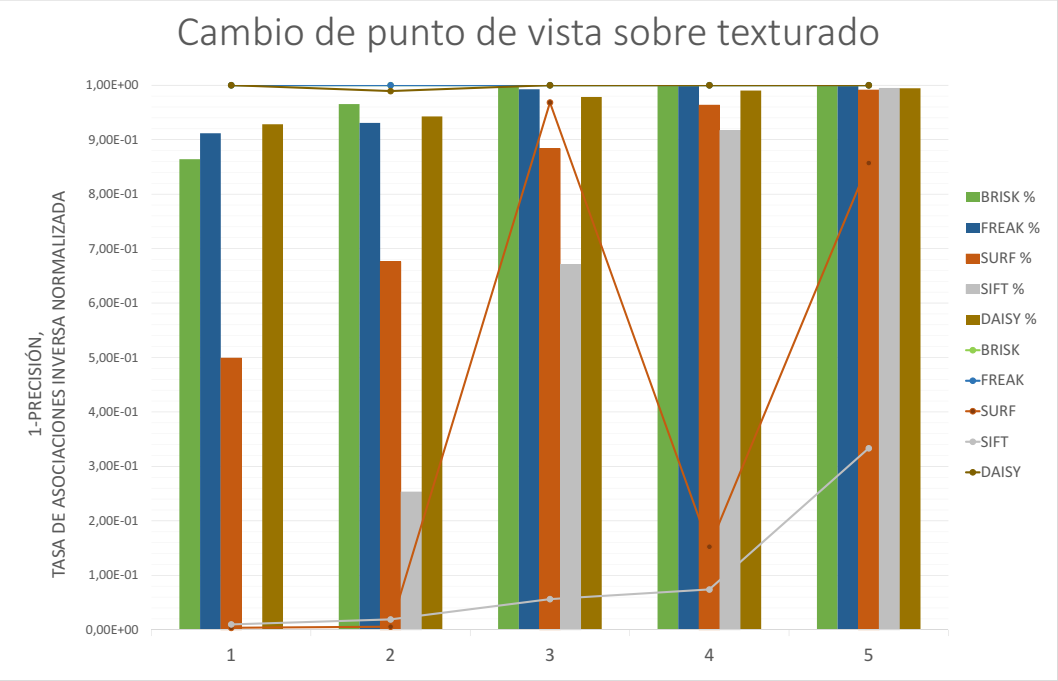


Figura 5.13: Resultados secuencia de 'Cambio de punto de vista sobre texturado' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.1.7. Cambio de Iluminación

Como se comentó en el capítulo anterior, los malos resultados del cuarto nivel de oscurecido, sólo pueden ser debidos a un mal cálculo de la homografía por parte de los autores. En el resto de niveles si que es posible observar como SIFT obtiene unos resultados excelentes para cambios de iluminación, seguido de SURF que consigue la misma precisión que SIFT pero con peor tasa de asociaciones. Entre los otros tres algoritmos, de nuevo el que consigue los peores resultados es DAISY.

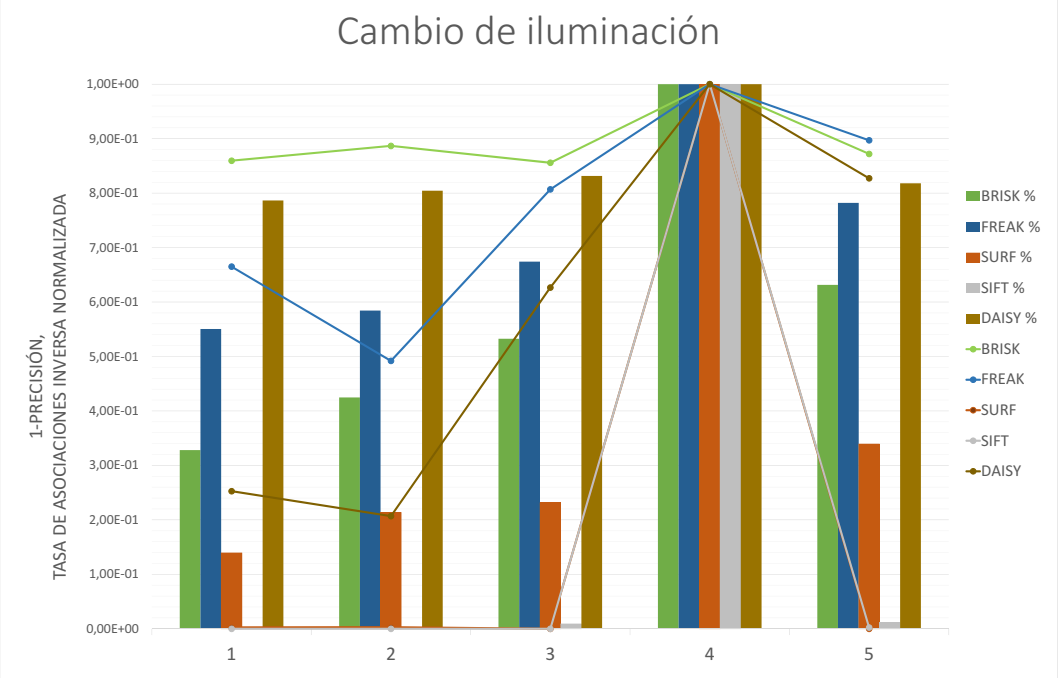


Figura 5.14: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.1.8. Compresión JPEG

De nuevo SIFT obtiene para esta secuencia unos resultados excelentes en comparación con el resto de algoritmos. En segundo lugar estaría SURF con una precisión muy alta igualmente pero con peores tasas de asociación. En tercer lugar estará BRISK con un nivel de precisión medio pero con peores tasas de asociación. Tanto DAISY como FREAK consiguen unas tasas de asociación muy pobres pero en el caso de FREAK además obtiene los peores resultados de precisión.

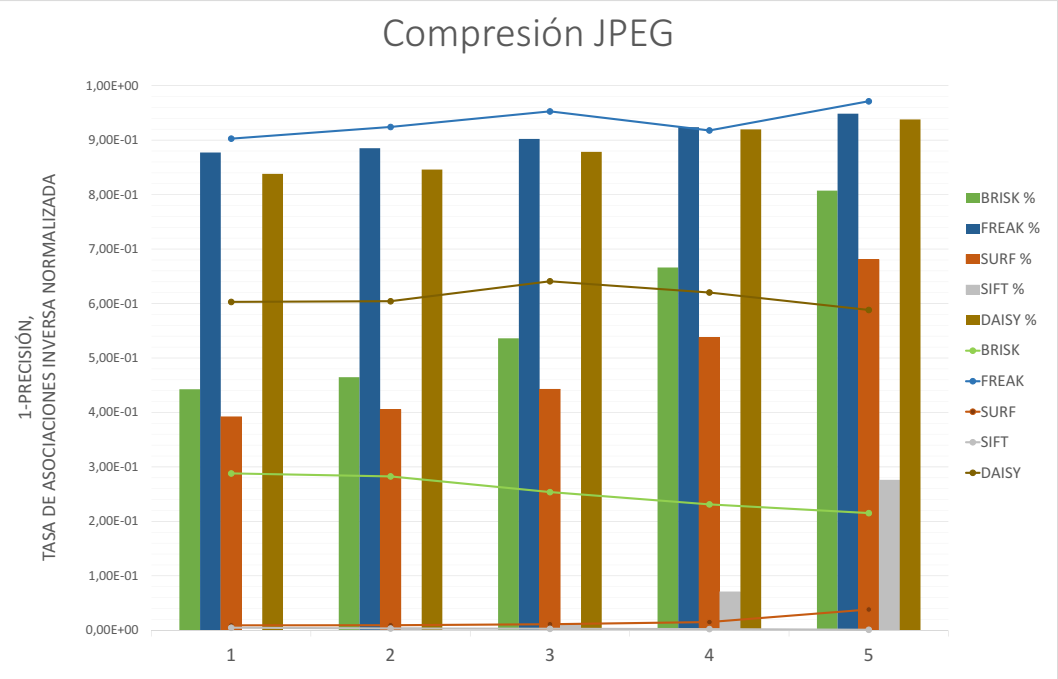


Figura 5.15: Resultados secuencia de 'Compresión JPEG' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.1.9. Conclusiones generales

Como se ha comentó en la evaluación de detectores, algunas secuencias del dataset presentan unas condiciones demasiado duras para todos los algoritmos que hacen que sea muy difícil extraer conclusiones a la vista de los datos.

Se observa también un segundo grupo de secuencias en las que todos los algoritmos salvo SIFT obtienen unos resultados demasiado pobres como para ser evaluados. Sólo es posible afirmar en estos casos que SIFT supera claramente a los demás algoritmos ante transformaciones como cambios de punto de vista y blurring sobre escenas

texturadas, pero entre el resto de los algoritmos es difícil establecer un ranking.

Por último, están las secuencias con las transformaciones de compresión JPEG, cambio de iluminación y blur, sobre las que si es posible ver una evolución a lo largo de los niveles de afectación para todos los algoritmos. En todos los casos aún así ha sido SIFT el que ha obtenido unos resultados claramente superiores al resto seguido de lejos por SURF.

A la vista de estos resultados, se refuerza la motivación del nuevo dataset aportado, debido a la imposibilidad de evaluar a algunos de los algoritmos sobre una buena parte de las transformaciones, a causa de los malos resultados que obtienen.

5.3.2. Evaluación sobre el nuevo conjunto de datos aportado.

5.3.2.1. Blur

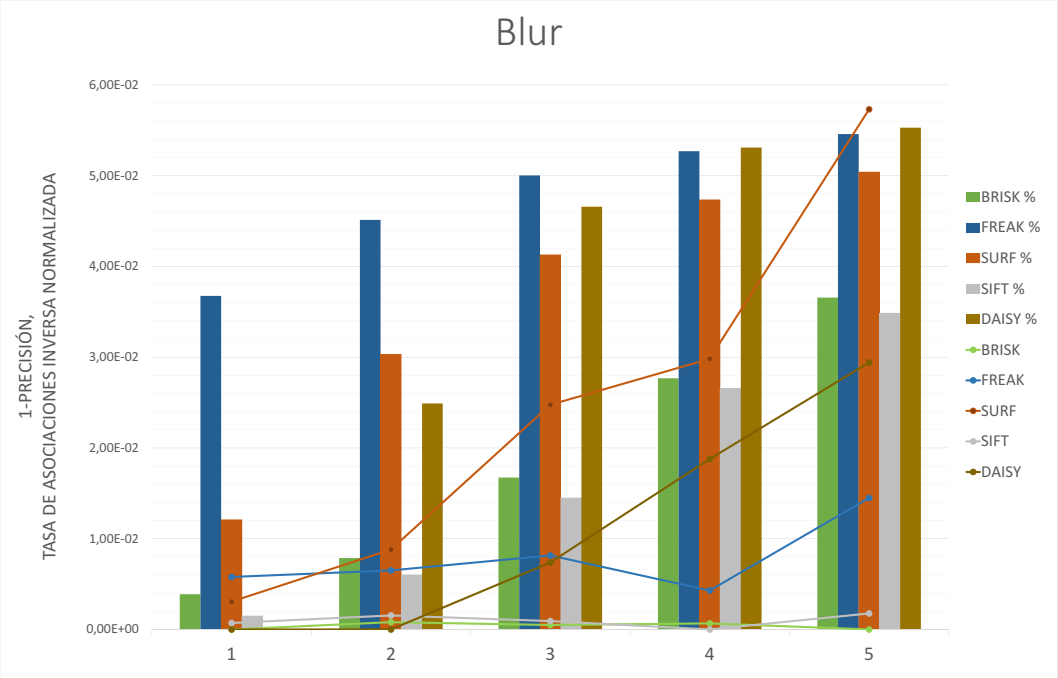


Figura 5.16: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

En ésta primera secuencia se puede ver cómo todos los algoritmos se mueven en un rango de precisión inversa que no supera el valor de 0,06, lo que es un muy buen resultado en todos los casos. Por otro lado, la tasa de asociaciones si que muestra una evolución claramente descendente. Los dos con mejores resultados en este caso son

SIFT y BRISK, siendo ligeramente mejores los de SIFT

5.3.2.2. Blur combinado con cambio de iluminación global

De nuevo se puede ver como SIFT y BRISK obtienen los mejores resultados, ganando SIFT en tasa de asociaciones y BRISK en precisión. En éste caso es SURF el que obtiene los peores resultados de precisión en los últimos pasos de afectación teniendo además una mala tasa de asociaciones.

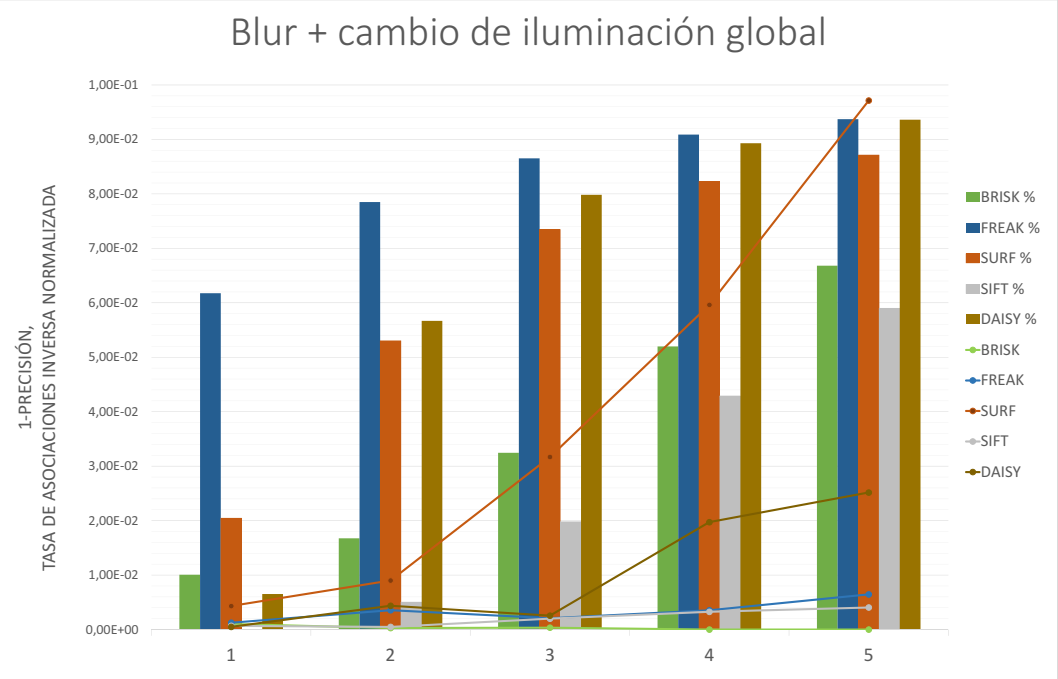


Figura 5.17: Resultados secuencia de 'Blur+cambio de iluminación global' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.3. Blur global con cambio de iluminación sobre un elemento

Se puede ver como el comportamiento de los algoritmos es muy similar al de la transformación global. Al igual que en la secuencia anterior, se observa que SURF es el peores resultados obtiene para este tipo de transformaciones mientras que SIFT obtiene las mejores valores de tasa de asociaciones y BRISK de precisión.

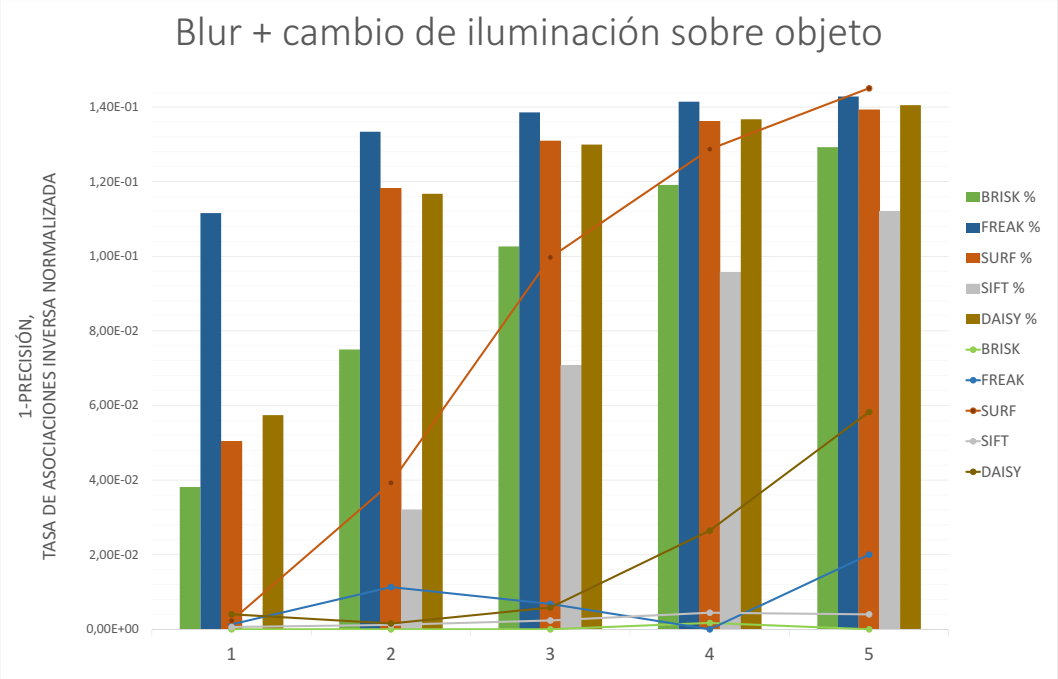


Figura 5.18: Resultados secuencia de 'Blur+cambio de iluminación sobre objeto' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.4. Blur global combinado con cambio de iluminación por áreas

En el caso de ésta secuencia que contiene un cambio de iluminación más irregular combinado con blurring y supone un mayor reto para los algoritmos que los anteriores vuelve a observarse como SIFT supera al resto de algoritmos claramente, tanto en precisión como en tasa de asociaciones. Es destacable también como BRISK empeora su rendimiento con respecto a las secuencias en las que el cambio de iluminación varía uniformemente por toda la imagen.

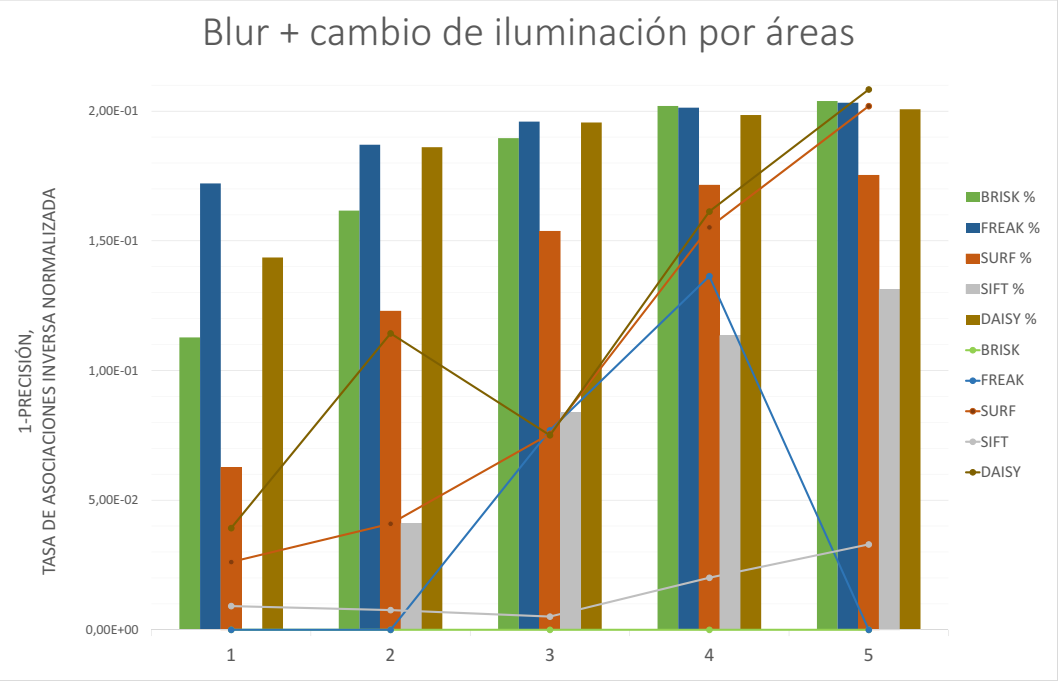


Figura 5.19: Resultados secuencia de 'Blur+cambio de iluminación por áreas' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.5. Blur combinado con cambio de punto de vista

Como se ha visto en secuencias anteriores, el blurring es una transformación que penaliza bastante a los algoritmos en cuanto a tasa de asociaciones. En este caso, esta combinación de cambio de punto de vista con blurring, ha sido la que ha obtenido unos resultados peores en comparación con otras combinaciones, por lo que se puede decir que el cambio de punto de vista también supondrá un reto para los algoritmos.

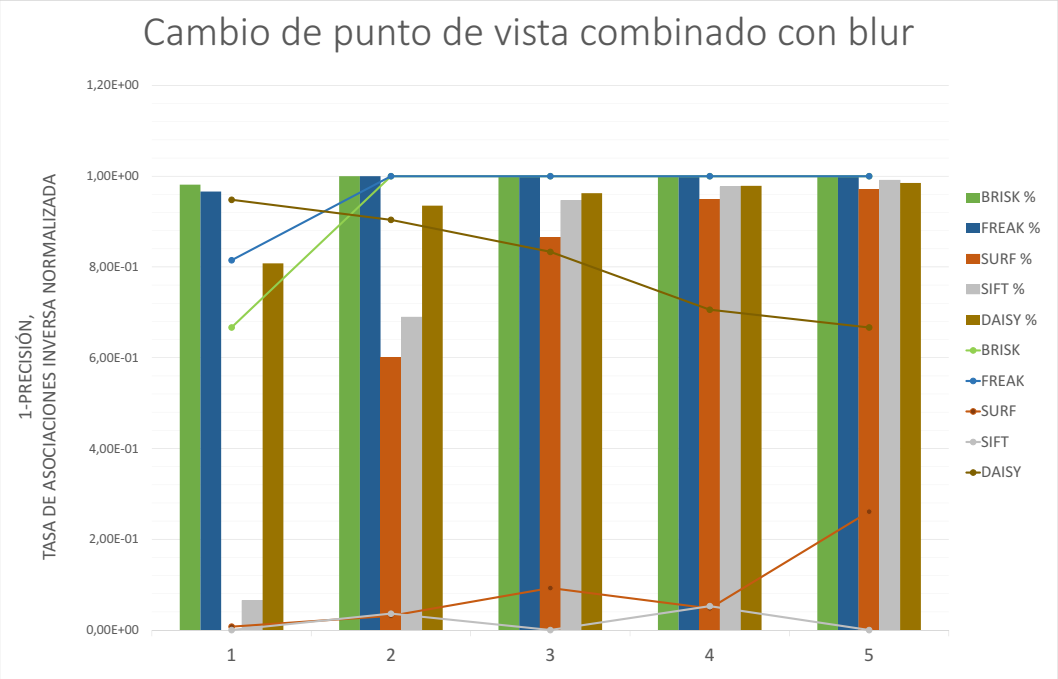


Figura 5.20: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.6. Cambio de iluminación global

A la vista de los buenos resultados tanto de precisión como de tasa de asociaciones de todos los algoritmos, se puede decir que el cambio de iluminación uniforme que modela la secuencia no supone una gran dificultad para éstos y habría que endurecer más los oscurecidos para poder observar una mayor variabilidad en los resultados.

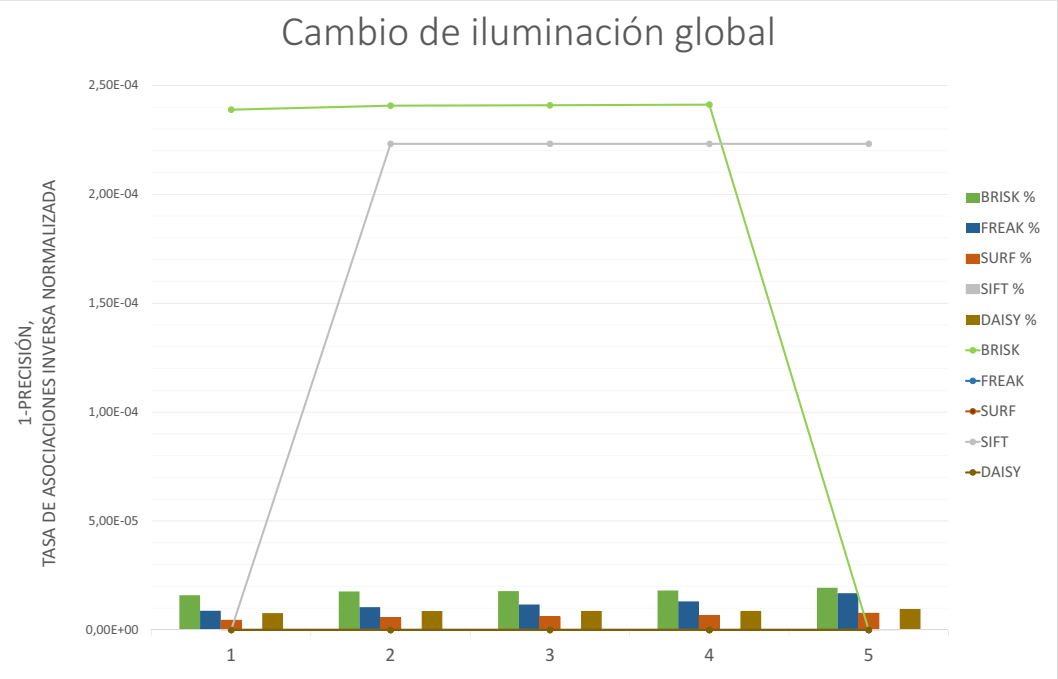


Figura 5.21: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.7. Cambio de iluminación sobre objeto

A diferencia de la secuencia anterior, aunque los resultados de precisión de los algoritmos siguen siendo muy buenos, la tasa de asociaciones si que se va reduciendo a lo largo de los pasos. Se puede observar una vez más como SIFT presenta los mejores valores de tasa de asociaciones y BRISK los peores.

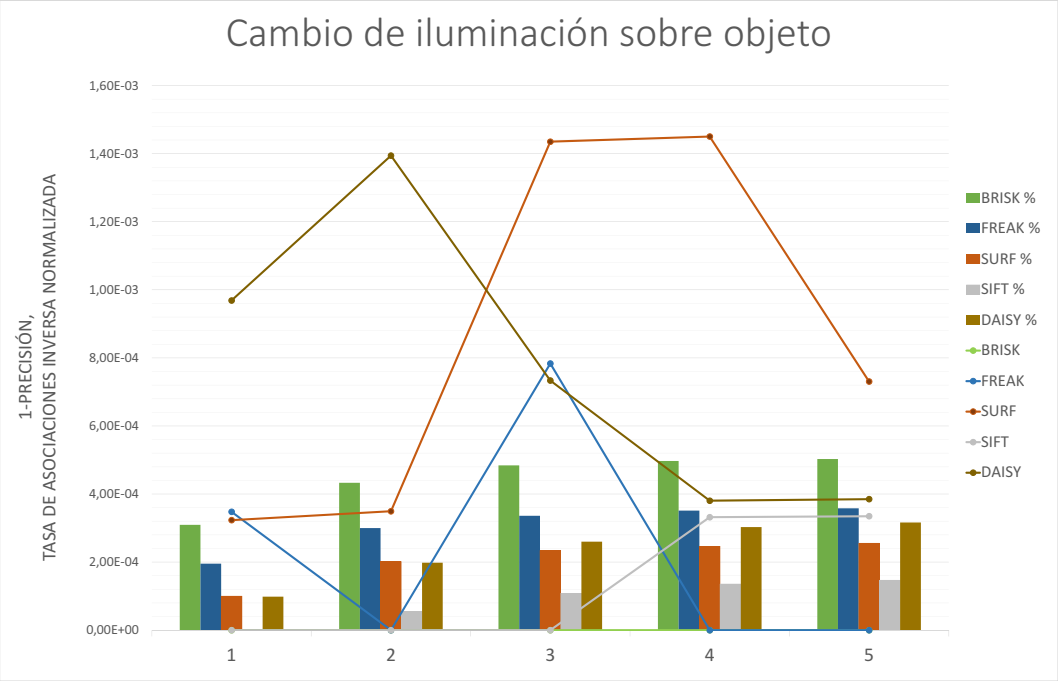


Figura 5.22: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.8. Blur lineal sobre objeto

Una vez más los valores de precisión de todos los algoritmos son bastante altos, siendo los de SURF los más bajos. En el caso de la tasa de asociaciones, SIFT vuelve a ser el mejor.

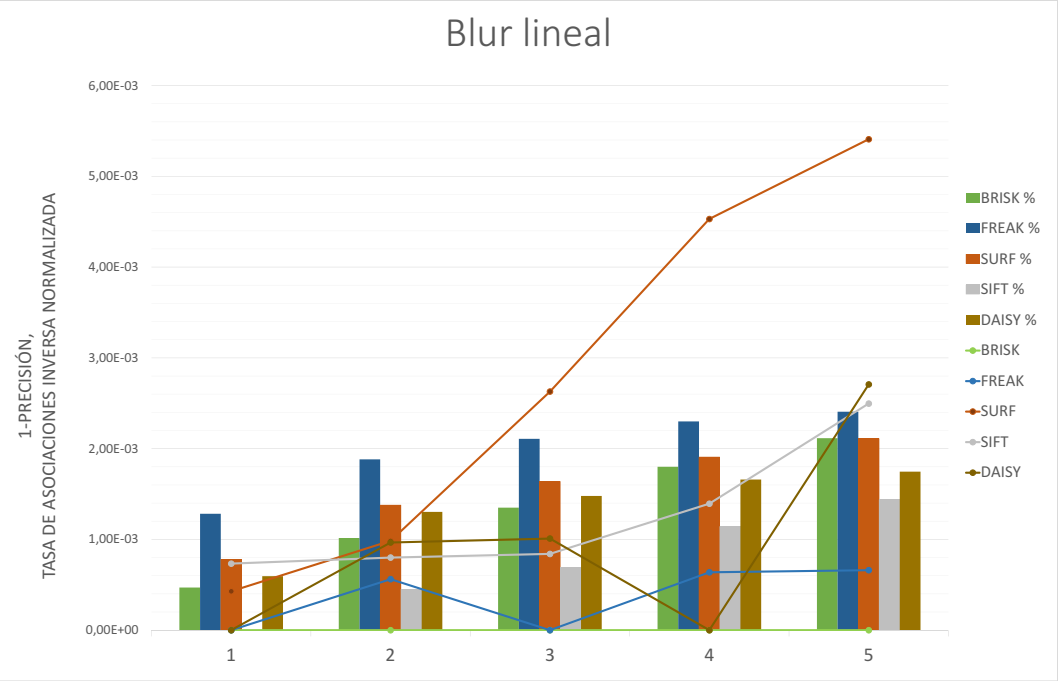


Figura 5.23: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.9. Escala combinado con rotación

Los resultados de precisión de SURF, SIFT y DAISY, indican que, pese a que la tasa de asociaciones si que se va viendo reducida en los sucesivos pasos, los puntos que logran describir son muy distintivos. Éste no es el caso de BRISK y FREAK que obtienen una precisión mucho menor así como también muestran peores resultados en tasa de asociaciones. En este caso BRISK es el que obtiene los mejores resultados, ligeramente por delante de SIFT y DAISY.

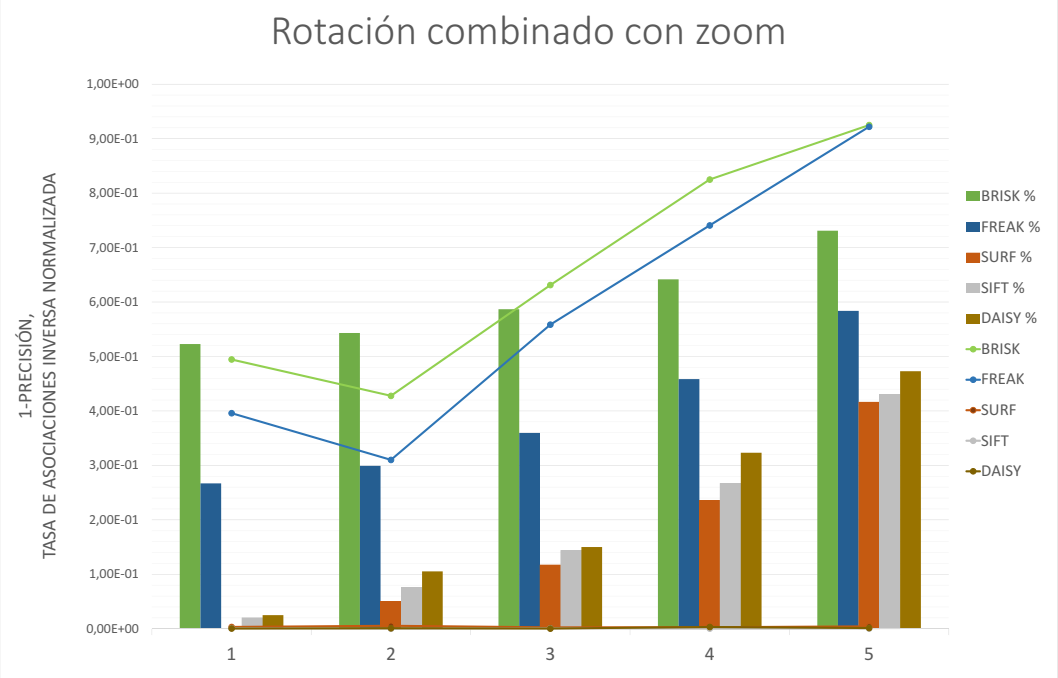


Figura 5.24: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.10. Ensombrecido con expansión en área

Como se ha comentado, los cambios de iluminación de ésta secuencia son muy irregulares. Los resultados de SIFT vuelven a sobresalir frente a los demás algoritmos, a pesar de que BRISK obtiene los mejores resultados de precisión, la tasa de asociaciones de SIFT es muy superior, por lo que se puede considerar que es el que mejores resultados obtiene.

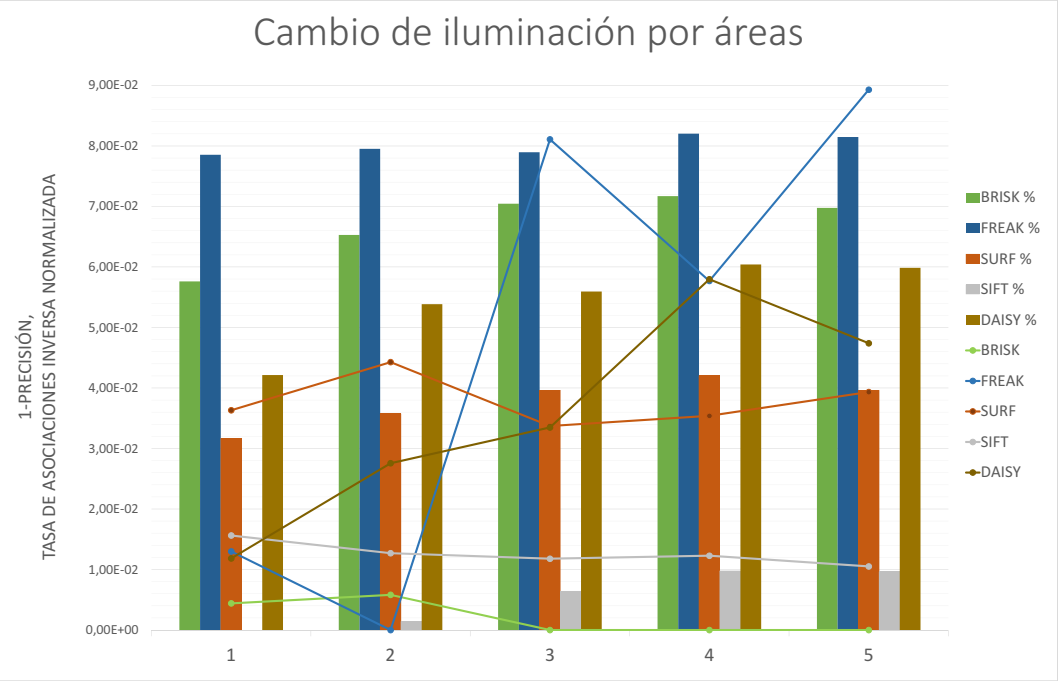


Figura 5.25: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.11. Cambio de punto de vista

Al igual que ocurría en la secuencia de rotación y escala, SURF es el algoritmo que obtiene los mejores resultados, en ésta ocasión con mayor diferencia respecto al SIFT en segundo lugar. Ambos obtienen una precisión muy alta, pero la tasa de asociaciones de SURF es mayor en todos los casos.

Es reseñable el aumento de precisión en las últimas imágenes de la secuencia de BRISK, FREAK, y DAISY, resultados que están derivados de la baja tasa de asociaciones, lo que refuerza, como se ha comentado, la necesidad de mostrar los valores de precisión y tasa de asociaciones en conjunto.

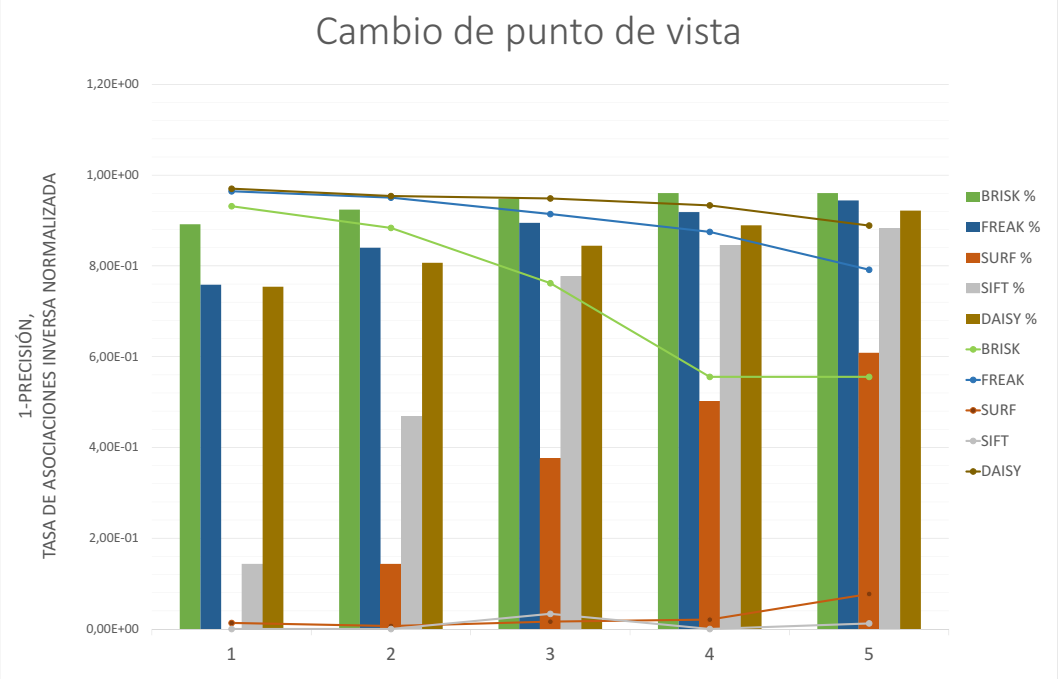


Figura 5.26: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.12. Cambio de punto de vista combinado con cambio de iluminación

Como se ha visto para otras combinaciones de cambio de punto de vista, es una transformación que supone un reto para todos los algoritmos. En este caso en combinación con cambios de iluminación resulta en una tasa de asociaciones muy baja de todos los algoritmos a partir del segundo nivel de afectación.

En el primer nivel de afectación resaltan los buenos resultados de SIFT, mientras que en los siguientes pasos es SURF el que obtiene unos resultados mejores que el resto de algoritmos.

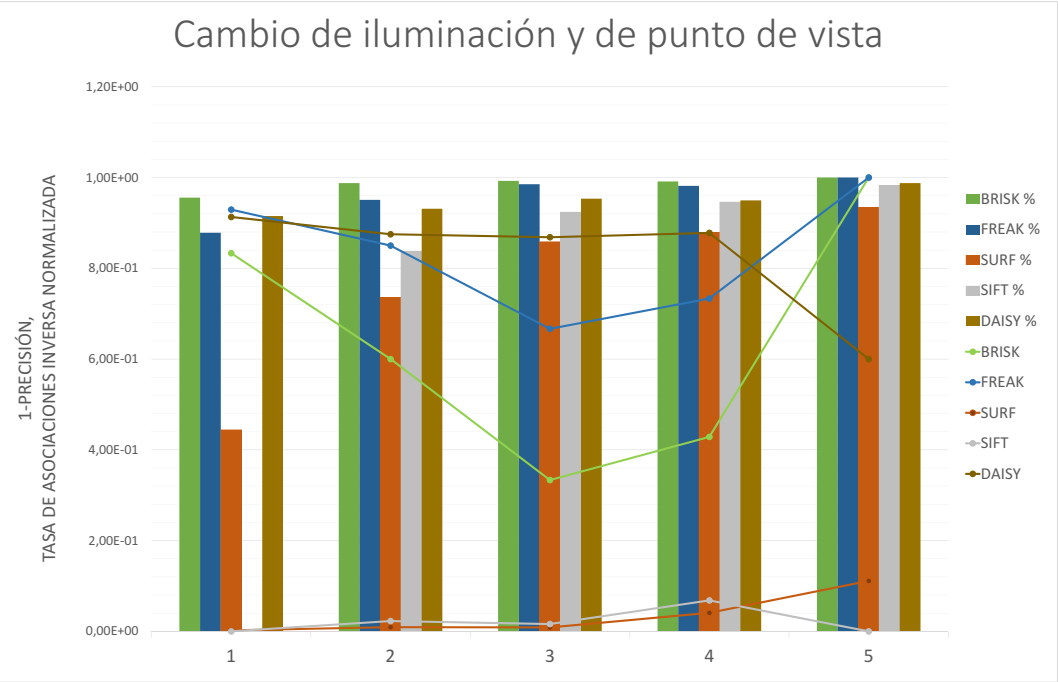


Figura 5.27: Resultados secuencia de 'Blur' para cada imagen de la secuencia con respecto a la primera: Cada línea representa el valor de 1-Precisión del algoritmo. Las columnas representan el porcentaje normalizado de asociaciones/correspondencias.

5.3.2.13. Conclusiones generales

Como se apuntó en la sección 5.3.1.9 el dataset aportado para este proyecto se ha construido evitando movimientos de cámara cuando la transformación no lo requiere. Gracias a esto se evita el paso de obtención de homografías en estas secuencias, lo que evita posibles fallos y además permite que en la evaluación se aíslen los efectos que se desea medir de otros efectos derivados de los movimientos de cámara.

Además se han evaluado los algoritmos sobre una variedad de nuevas transformaciones, aumentando la complejidad de las mismas y añadiendo nuevas combinaciones,

lo que ha aportado una mayor riqueza a la evaluación al poder extraer nuevas conclusiones en base a los comportamientos de los algoritmos en estas situaciones.

5.4. Conclusiones.

Tanto en la evaluación sobre el dataset de referencia como sobre el nuevo dataset construido, los algoritmos que mejores resultados han obtenido han sido los populares SIFT y SURF.

SIFT ha sido el algoritmo que ha mostrado una mayor robustez frente a las transformaciones de tipo cambio de iluminación y blurring.

Por otro lado SURF ha sido, claramente, el que mejores resultados ha obtenido frente a cambios de escala y rotación así como de cambio de punto de vista.

La gran desventaja de SIFT con respecto al resto de algoritmos está en algo que no se ha cuantificado en ésta evaluación, que es el tiempo de ejecución. Mientras que algoritmos como DAISY, BRISK y FREAK pueden funcionar a tiempo real, SIFT está muy lejos de estos tiempos de ejecución.

La comparación con la evaluación teórica quedaría de la siguiente manera:

En el caso de SIFT confirma la valoración inicial en cuanto a cambios de iluminación y blur, mientras que SURF mejora las expectativas frente a cambios de punto de vista, escala y rotación.

En los casos de DAISY, FREAK y BRISK, todos ellos empeoran en algún sentido las puntuaciones iniciales si se les compara con los resultados de SURF y SIFT.

Capítulo 6

Conclusiones y trabajo futuro.

6.1. Conclusiones.

Como lectura global del proyecto, se ha aportado un marco de evaluación sobre el que se han evaluado once algoritmos de detección y cinco algoritmos de descripción del estado del arte.

En cuanto a los diferentes objetivos parciales que se habían planteado, se pueden extraer una serie de conclusiones.

En base al estudio del estado del arte realizado, se han propuesto dos clasificaciones de los diferentes algoritmos de detección y descripción. Esta categorización ha permitido actualizar los algoritmos considerados más punteros del estado del arte.

Analizando los algoritmos seleccionados, se puede concluir que en el estado del arte de la detección comienza a trabajarse con mayor intensidad en las propuestas basadas en técnicas de descripción de entorno dado su menor coste computacional. Por su parte, en el estado del arte de la descripción comienzan a incrementarse las técnicas de descripción binarias, de nuevo por un motivo principal de coste computacional.

Se han detectado defectos que presentaban anteriores marcos de evaluación, como era la presencia de secuencias con poca utilidad o con tasas de acierto tan bajas que impedían extraer conclusiones, o secuencias con bajas resoluciones. El marco de evaluación propuesto por su parte, ha conseguido recoger una mayor riqueza de propiedades a evaluar, incrementando el número de secuencias, su calidad y su utilidad de cara a la interpretación del funcionamiento de los algoritmos en diferentes condiciones.

La evaluación de técnicas de detección arroja que se trata de una tarea profundamente trabajada. Estudiada de forma independiente a la de descripción favorece a aquellas ténicas que se dedican únicamente a esta primera etapa frente a las que incorporan detección y descripción en su propuesta. Una excepción es la técnica de

SURF, que pese a tener un mayor coste computacional que las técnicas de detección de entorno, ha resultado se de la de mayor fiabilidad. En cuanto a la línea de trabajo que se puede extraer de este análisis, parece que las evoluciones de la técnica de AGAST son las más probables dado que presenta grandes resultados pese a su relativa novedad.

Por último, en la evaluación de las técnicas de descripción se puede observar la tendencia del estado del arte hacia los descriptores binarios que mejoran notablemente los costes computacionales de las técnicas tradicionales. Se han presentado dos propuestas como referentes, FREAK y BRISK, esta última con unos resultados cuantitativos cercanos a los de las técnicas más fiables del estado del arte. Atendiendo únicamente a los resultados cuantitativos, el tradicional descriptor de SIFT ha obtenido mejor resultados que el resto de los algoritmos en la mayoría de las secuencias evaluadas. Sin embargo el coste computacional de éste algoritmo es mucho mayor que el del resto, lo que hace que no sea posible su uso en aplicaciones que requieran una mayor eficiencia.

6.2. Trabajo futuro.

Tras la realización de éste proyecto, se consideran dos líneas principales de trabajo futuro.

La primera de ellas es relacionada con el propio proyecto, y con el marco de evaluación más concretamente. En está línea, los principales focos de trabajo considerados serían:

- Realizar un estudio de las distintas aplicaciones del estado del arte, observando cuáles serían las propiedades de los puntos de interés más relevantes según cada una. En base a esto, se podría enriquecer la evaluación aportando datos en cuanto a que técnicas de puntos serían las idóneas para según que aplicación.

- Incrementar el conjunto de datos con nuevas propiedades, que se podrían seleccionar en base al estudio orientado a aplicaciones mencionado.

- Incluir métricas para la evaluación del coste computacional de las diferentes técnicas.

La segunda de ellas, ya referida a trabajo a partir de la conclusión del proyecto, presenta innumerable posibilidades. A continuación se detallarán las dos que se han considerado más fructíferas.

- Desarrollo de una herramienta de evaluación y ranking *online* de algoritmos que utilice como base de evaluación el marco propuesto en este proyecto y los algoritmos ya evaluados.
- Desarrollo de una herramienta de apoyo a la investigación, que permita sugerir al usuario que técnicas emplear para según que tarea. Se podría incluso aportar secuencias anotadas a la herramienta, y que ésta, en base a una evaluación de técnicas sobre la misma sugiriese que algoritmo presenta mayores posibilidades de éxito.

Bibliografía

[1] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, pp. 404–417, Springer, 2006.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[3] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European conference on computer vision*, pp. 430–443, Springer, 2006.

[4] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *European conference on Computer vision*, pp. 183–196, Springer, 2010.

[5] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *2011 International conference on computer vision*, pp. 2548–2555, IEEE, 2011.

[6] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 5, pp. 815–830, 2010.

[7] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Computer vision and pattern recognition (CVPR), 2012 IEEE conference on*, pp. 510–517, Ieee, 2012.

[8] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, L. Cehovin, G. Fernandez, T. Vojir, G. Hager, G. Nebehay, and R. Pflugfelder, "The visual object tracking vot2015 challenge results," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1–23, 2015.

[9] F. Attneave, "Some informational aspects of visual perception.," *Psychological review*, vol. 61, no. 3, p. 183, 1954.

[10] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and trends® in computer graphics and vision*, vol. 3, no. 3, pp. 177–280, 2008.

[11] H. P. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover.," tech. rep., DTIC Document, 1980.

[12] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Alvey vision conference*, vol. 15, p. 50, Citeseer, 1988.

[13] P. Sankar and C. Sharma, "A parallel procedure for the detection of dominant points on a digital curve," *Computer Graphics and Image Processing*, vol. 7, no. 3, pp. 403–412, 1978.

[14] J. L. Crowley and A. C. Sanderson, "Multiple resolution representation and probabilistic matching of 2-d gray-scale shape," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 113–121, 1987.

[15] T. Lindeberg, "Scale-space for discrete signals," *IEEE transactions on pattern analysis and machine intelligence*, vol. 12, no. 3, pp. 234–254, 1990.

[16] T. Lindeberg, "Feature detection with automatic scale selection," *International journal of computer vision*, vol. 30, no. 2, pp. 79–116, 1998.

[17] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International conference on computer vision*, pp. 2564–2571, IEEE, 2011.

[18] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[19] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63–86, 2004.

[20] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *European conference on computer vision*, pp. 128–142, Springer, 2002.

[21] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[22] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, Ieee, 1999.

[23] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[24] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II–506, IEEE, 2004.

[25] S. Lazebnik, C. Schmid, and J. Ponce, "Sparse texture representations using affine-invariant neighborhoods," in *In Proc. IEEE Conf. Comp. Vision Patt. Recog*, Citeseer, 2003.

[26] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[27] J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system," *Biological cybernetics*, vol. 55, no. 6, pp. 367–375, 1987.

[28] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or how do i organize my holiday snaps?," in *European conference on computer vision*, pp. 414–431, Springer, 2002.

[29] L. Van Gool, T. Moons, and D. Ungureanu, "Affine/photometric invariants for planar intensity patterns," in *European Conference on Computer Vision*, pp. 642–651, Springer, 1996.

[30] A. Canclini, M. Cesana, A. Redondi, M. Tagliasacchi, J. Ascenso, and R. Cilla, "Evaluation of low-complexity visual feature detectors and descriptors," in *Digital Signal Processing (DSP), 2013 18th International Conference on*, pp. 1–7, IEEE, 2013.

[31] O. Miksik and K. Mikolajczyk, "Evaluation of local detectors and descriptors for fast feature matching," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 2681–2684, IEEE, 2012.

[32] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *European conference on computer vision*, pp. 778–792, Springer, 2010.

[33] B. Fan, F. Wu, and Z. Hu, "Rotationally invariant descriptors using intensity order pooling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2031–2045, 2012.

[34] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *2011 International Conference on Computer Vision*, pp. 603–610, IEEE, 2011.

[35] J. Heinly, E. Dunn, and J.-M. Frahm, "Comparative evaluation of binary features," in *Computer Vision–ECCV 2012*, pp. 759–773, Springer, 2012.

[36] F. Remondino, "Detectors and descriptors for photogrammetric applications," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 3, pp. 49–54, 2006.

[37] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual slam," *Machine Vision and Applications*, vol. 21, no. 6, pp. 905–920, 2010.

[38] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263–284, 2007.

[39] F. Tombari, S. Salti, and L. Di Stefano, "Performance evaluation of 3d keypoint detectors," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 198–220, 2013.

[40] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[41] G. Dorkó and C. Schmid, "Selection of scale-invariant parts for object class recognition," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 634–639, IEEE, 2003.

[42] B. Leibe and B. Schiele, "Scale-invariant object categorization using a scale-adaptive mean-shift search," in *Joint Pattern Recognition Symposium*, pp. 145–153, Springer, 2004.

[43] J. Liu, J. Luo, and M. Shah, "Recognizing realistic actions from videos in the wild," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1996–2003, IEEE, 2009.

[44] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of computer vision*, vol. 37, no. 2, pp. 151–172, 2000.

[45] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[46] T. Lindeberg, "Image matching using generalized scale-space interest points," in *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 355–367, Springer, 2013.

[47] S. M. Smith and J. M. Brady, "Susan - a new approach to low level image processing," *International journal of computer vision*, vol. 23, no. 1, pp. 45–78, 1997.

Apéndice A

Presupuesto

1. **Ejecución Material**
   - Compra de ordenador personal (Software incluido) ..... 2.000€
   - Alquiler de impresora láser durante 6 meses ..... 260 €
   - Material de oficina ..... 150 €
   - Total de ejecución material ..... 2.400 €
2. **Gastos generales**
   - 16 % sobre Ejecución Material ..... 352 €
3. **Beneficio Industrial**
   - 6 % sobre Ejecución Material ..... 132 €
4. **Honorarios Proyecto**
   - 1800 horas a 15 € / hora ..... 27.000 €
5. **Material fungible**
   - Gastos de impresión ..... 280 €
   - Encuadernación ..... 200 €
6. **Subtotal del presupuesto**
   - Subtotal Presupuesto ..... 32.774 €

7. **I.V.A. aplicable**

   - 21 % Subtotal Presupuesto ..... 6.882,5 €

8. **Total presupuesto**

   - Total Presupuesto ..... 39.656,5 €

Madrid, julio 2016

El Ingeniero Jefe de Proyecto

Fdo.: Miguel Martín Redondo

Ingeniero de Telecomunicación

Apéndice B

Pliego de condiciones

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de una evaluación comparativa de detectores y descriptores de puntos de interés en imágenes. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicha evaluación. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego. Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado

en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4 % del presupuesto y la provisional del 2 %.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean

oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.